Tracking buffers of mutation and noise to the genome

Jennifer Lachowiec

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Christine Queitsch, Chair

Elhanan Borenstein

Maitreya Dunham

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Tracking buffers of mutation and noise to the genome

Jennifer Lachowiec

Chair of the Supervisory Committee:
Associate Professor Dr. Christine Queitsch
Department of Genome Sciences

Phenotypes are buffered from both genetic perturbations and developmental noise; however, the mechanisms by which this buffering occurs and its evolutionary relevance are poorly understood. In this dissertation, a combination of genetic and computational approaches were undertaken to not only identify genetic loci that are buffered from phenotypes but also map genes that provide phenotypic buffering. For example, I found that substrates of Heat shock protein 90 (Hsp90), the best-understood source of buffering, tend to accumulate genetic changes in a manner that affects evolution. Hsp90 is also a proven buffer of developmental noise, so the mechanism by which this ability arises was explored. To expand the number of known developmental noise buffers, innovative methods for genome-wide association were used to map novel regulators. The ability to identify loci that are buffered and provide buffering indicates that the distribution of phenotypes across a population arises through complex interactions between genetic loci, including genes that act as buffers.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I wish to offer my sincere thanks to:

Christine Queitsch, my advisor, whose ceaseless creativity inspires me to look beyond the obvious and whose belief in me made every day of my PhD a pleasure.

Members of the Queitsch lab, present and past for their friendship, guidance, and intellectual insights (and help with experiments): Keisha Carlson, Karla Schultz, Sanna Sullivan, Max Press, Michael Dorrity, Cris Alexandre, Tzitziki Lemus, Pauline Rival, Kerry Bubb, Alex Mason, Beth Morton, James Urton Janne Lempe, Soledad Undurraga, Jacob Bale, Hannah Jordt, and Anukul Veeraraghavan.

Jennifer Nemhauser, who has been an especial mentor in career planning and the grad school process.

Patrick Murphy and Örjan Carlborg. who both welcomed me into their labs to learn techniques and for invaluable scientific discussion.

My committee: Jennifer Nemhauser, Maitreya Dunham, Elhanan Borenstein, Keiko Torii, and Doug Fowler for your input and encouragement as my project transformed over the years.

My MCB cohort and Genome Sciences graduate students for sharing the process that is a PhD.

Michelle and Chris for sharing many duck dinners, bike rides, and the trials of living a ferry ride away.

The WPCGA for being a warm and welcoming group of people that can understand why I don’t want my dogs inbred.

My family: my mom and my dad who supported every single one of my endeavors without question, always believing that I have made the right choice, and my brother for whom I aspire to inspire.

Zeb, whose belief in me and my goals surpasses my own.

My dogs Benny and Bam: need I say more?

CHAPTER ONE

INTRODUCTION[1]

Variation in phenotypes is crucial to the process of evolution, allowing natural selection to promote the sub-range of a phenotypic distribution that is most adaptive (Darwin 1859). Across a population of individuals, the distribution of each phenotype is described by both a measure of its center and dispersion, the mean and variance, respectively. This distribution arises from each individual's phenotypes, which are determined by three factors: genetic background, environments encountered, and error or noise (Hartl & Clark 2007). These three components affect the mean and variance of phenotypes both additively and through their interactions. Much success has been achieved in determining the additive effects of genetics and environment on phenotypes, especially in cases where a phenotype is controlled by a small number of genes (Queitsch et al. 2012). Efforts to understand how interactions among genetic background, environment, and noise muddy the translation of genotype to phenotype are beginning in earnest.

The complexity of the interactions among genetics, environment, and noise partially explains the difficulty in identifying each genetic variant that influences phenotypes (Queitsch et al. 2012). First, there are interactions within the genetic component of phenotypic variation. Genetic variation at one locus can influence the effect of another locus on a phenotype (Figure 1.1a,b). The number of interactions including one specific genetic locus can vary, ranging from a

[1] Portions of this chapter are adapted from "Molecular mechanisms of robustness in plants," Current Opinion in Plant Biology: 16(1), 2013, pp. 62-69, by J. Lempe, J. Lachowiec, A. M. Sullivan and C. Queitsch.

locus that interacts with only one other locus, to strong genetic modifiers that interact with many other loci (Lempe et al. 2012). Second, genetic variation can interact with noise (Figure 1.1c). Noise can be traced to stochastic errors in development due to the diffusion of low concentrations of small molecules within cells (Masel & Siegal 2009). The degree of developmental noise in a trait is measured by determining the variance of that trait when controlling both the environment and genetic background. This measured variance can be a genetically controlled quantitative trait (Hall et al. 2007). Finally, the environmental component of phenotypic variation can also interact with the genetic component. In other words, different genetic backgrounds may have different reaction norms. Notably, there is some evidence that the same gene can interact with all three of these components, genetic background, noise and environment (Queitsch et al. 2002; Lehner 2010).

From a biochemical point of view, likely candidates for strong genetic modifiers are genes that encode proteins with ubiquitous cellular functions. Hsp90 is one such gene that encodes a protein with hundreds of proven protein interactions (Zhao et al. 2005; Taipale et al. 2012). Indeed, studies have mapped loci for dozens of traits that are Hsp90-dependent (Todd A Sangster et al. 2008; T Sangster et al. 2008; Jarosz & Lindquist 2010). Hsp90 also affects the variance associated with quantitative traits, indicating that Hsp90 also regulates developmental noise (Queitsch et al. 2002). Little progress has been made thus far to understand how developmental noise systematically affects the distribution of phenotypes, though the potential influence of developmental noise has long been appreciated (Waddington 1942).

It is evident that Hsp90 impacts both the mean and variance of phenotypic distributions through its interaction with other loci and developmental noise. Whether Hsp90's

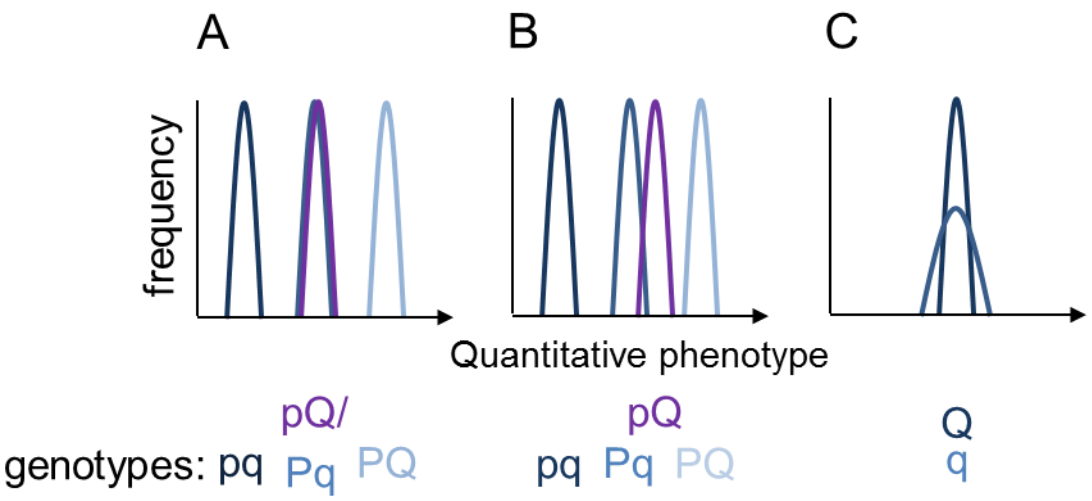

Figure 1.1. Effects of interactions on quantitative traits. A & B) Pictured are two different types of genetic interactions. A) Two loci (p/P and q/Q) with equal allelic effects (P and Q) interact additively in this phenotype. B) Two loci interact epistatically to regulate the phenotype. The Q allele alone has a greater effect than the P allele alone, but in combination, the Q allele has a lesser effect. C) Pictured is the interaction between genetics and developmental noise. The Q allele results in a small variance in the trait, while the presence of the q allele is associated with a larger variance in the trait in otherwise isogenic individuals in the same environment.

interaction with other genetic loci has influenced evolutionary processes is still unclear. Here, I present my efforts to detect Hsp90's interaction with other loci and the evolutionary effects of this interaction. I also probe Hsp90's interaction with developmental noise as measured by trait variance. Trait variance is crucial for natural selection to act, but its regulation is little studied. I explore the mechanisms by which developmental noise is regulated, including how Hsp90 maintains low levels of developmental noise.

Hsp90 is an essential eukaryotic protein chaperone

Hsp90 is highly conserved across eukaryotes, and its homologs are found in bacteria and archaea as well (Press et al. 2013; Bogumil et al. 2014). Hsp90 is essential across all eukaryotes tested and is ubiquitous, found not only in the cytoplasm, but also in most organelles and the extracellular matrix. Mitochondria- and chloroplast-specific isoforms of Hsp90s diverged early in the evolution of the Hsp90 family (Taipale et al. 2010), validating its important role in organelles. At ambient temperatures in *Saccharomyces cerevisiae*, Hsp90 protein levels can be reduced by 10-fold with no growth defects (Borkovich et al. 1989), indicating a reservoir of Hsp90 function beyond what is necessary for cellular function in physiological conditions.

Protein chaperones, such as Hsp90, were originally identified due to their up-regulated protein levels when temperatures were increased and were named thus for their respective migration position on protein gels (small Hsps, 60, 70, 90, 100). Protein chaperones facilitate the proper folding and maturation of proteins and are essential to maintain protein homeostasis. Most proteins are marginally thermodynamically stable in their structures, unfolding as temperatures increase (Hartl et al. 2011), and this loss of structure can be salvaged by chaperones. In fact,

protein chaperones respond to any stress that alters protein folding and stability, which includes changes to oxidative or ionic state or the presence of UV or heavy metals (Stensløkken et al. 2010; Park et al. 2010; Cid et al. 2010; Pratt et al. 2010; Sekimoto et al. 2010; Gerspacher et al. 2009). In addition to these roles in stresses, chaperones are essential in ambient conditions to maintain protein homeostasis. Large multi-domained proteins can take hours to fold or even never reach their native state *in vitro*, so within a cell, chaperones play a crucial role in facilitating folding of such proteins. Chaperones also aid proteins with structures that tend to misfold, and guide mis-translated proteins to the proteasome for degradation (Hartl et al. 2011). In addition, chaperones help proteins to cross membranes, unwinding proteins to pass through narrow transporters and re-fold at their destinations (Schleiff & Becker 2011). Protein chaperones also aid in signal transduction, priming proteins for post-translational modification and aiding in oligomerization (Zou et al. 1998; Gangaraju et al. 2011). All of these many functions are subdivided among the chaperones.

Chaperones are diverse in function and have a range of specificities for substrate proteins. For example, Hsp70 recognizes and binds short regions of exposed hydrophobic residues of unfolded polypeptides as they emerge from the ribosome. This polypeptide folds upon release from Hsp70. If the native state of the polypeptide is not obtained after release from Hsp70, Hsp70 will rebind and the process will cycle repeatedly until the native state is achieved. In addition to folding at the ribosome, Hsp70 is found at translocons and helps transport proteins across membranes. Moreover, as one of the most diversely acting chaperones, Hsp70 also acts on mature proteins and can target mis-folded proteins for degradation (Kampinga & Craig 2010). In contrast, Hsp60 does not have the diversity of function of Hsp70, though it does interact with a

large number of substrates. Hsp60, also known as the chaperonin or GroEL/ES in bacteria, forms a barrel (Xu et al. 1997) into which proteins can enter and fold in a protected environment (Clare et al. 2012). GroEL/ES reliance varies across proteins, ranging from complete dependence to complete independence (Kerner et al. 2005; Fujiwara et al. 2010). In addition to these chaperones that act initially in protein folding, several haperones also act further downstream in protein homeostasis. Similar to Hsp60, Hsp100 and the sHsps have one function, but can act on many substrates. Small Hsps act as holdases to prevent the aggregation of proteins, but in the event that proteins do mis-fold and form aggregates, Hsp100 disaggregates these complexes to prevent their toxic effects. (Hartl et al. 2011). Finally, some chaperones have specific functions in the assembly of protein complexes after contact with more generally acting chaperones, such as $RbcX_2$, the Rubisco chaperone (Liu et al. 2010).The diversity and necessity of chaperone functions may have arisen early in evolution to deal with the aggregation-prone environment of protein-packed cells (Hartl et al. 2011).

Hsp90 is a chaperone that functions later but at several key time-points in the life cycle of its protein substrates, receiving its partially folded substrates after their interaction with Hsp70. Hsp90 works as a dimer, cycling from a more open conformation to a more closed conformation through the hydrolysis of ATP. The transition between closed and open states promotes binding of Hsp90s substrates to the dimer; however why this structural cycle promotes substrate folding is still unclear (Hartl et al. 2011). Hsp90 interacts with at least 20 different proteins that regulate the kinetics of its ATP hydrolysis, recruit clients, and assist in client release (Röhl et al. 2013). In addition to Hsp90's role in substrate folding, interaction with Hsp90 is necessary for proper post-translational modifications (Zou et al. 1998; Gangaraju et al. 2011) and transfer across

membranes (Young et al. 2003). Hsp90 is also found in association with ubiquitin ligases (Taipale et al. 2012), suggesting that Hsp90 facilitates protein degradation (Murata et al. 2001; Morishima et al. 2008). Based on the myriad of functions of Hsp90, it is easy to see why it is essential to eukaryotes.

The substrates, or clients, of Hsp90 are numerous but select. Many of the over 300 Hsp90 clients were identified in studies of single proteins, though there have been efforts to identify Hsp90 physical and genetic interactors *en masse* via proteomics and synthetic genetic interaction screens (Zhao et al. 2005; McClellan et al. 2007). Many Hsp90 clients are involved in signal transduction, including protein kinases and transcription factors, such as the mammalian steroid receptors and the heat shock transcription factor HSF1 (Taipale et al. 2010). Unlike Hsp70, Hsp90 does not recognize particular amino acid residues. Comparisons of the sequences features among known Hsp90 clients have not detected any motifs or enrichments (Taipale et al. 2010; Röhl et al. 2013). Recently a systematic, high-throughput study tested over 300 human protein kinases for their interaction with Hsp90, discovering that two-thirds interact with Hsp90. Interestingly, although all mammalian protein kinases are related, client status is not determined by phylogeny. Sequence comparisons and domain swaps between kinases suggested that both localized and global features could regulate client status. The strongest support for a feature that defines client status was thermodynamic stability (Taipale et al. 2012). The hypothesis that Hsp90 recognizes metastability is supported by the finding that one amino acid change can greatly increase Hsp90 dependency (Citri et al. 2006) as most amino acid changes are destabilizing to a protein (DePristo et al. 2005). Based on these findings, clients cannot be predicted computationally unless a crystal structure is available. These data also suggest that

client status would be dynamic as a proteins evolve, indicating shifting dependency on Hsp90's chaperone activity.

The surface of Hsp90 that recognizes clients has been difficult to pinpoint, likely due to the diversity of the proteins with which Hsp90 interacts. Hsp90 is made up of three domains: N-terminal domain, middle domain, and C-terminal domain. Whereas all three domains have been implicated in client binding, the middle domain is most frequently found to bind clients (Röhl et al. 2013). For example, the protein tau, which abnormally accumulates in the brains of patients with Alzheimer's, is an Hsp90 client. Partially unfolded tau makes loose contacts with Hsp90 over a large surface, binding both the N-terminal and middle domains of Hsp90. It is becoming clearer that Hsp90 does indeed bind metastable, partially folded intermediates (Karagöz et al. 2014), perhaps explaining the broad substrate interaction surface of Hsp90.

While Hsp90 is highly conserved across eukaryotes, it varies within species in both copy number and degree of polymorphism. For example, while the *Caenorhabditis elegans* genome codes for one cytoplasmic copy of Hsp90, 13 cytoplasmic copies are found in the genome of *Glycine max* (Goodstein et al. 2012). Among bacteria, zero to three copies of the Hsp90 homolog hptG have been found (Press et al. 2013). Furthermore, the degree of polymorphism within Hsp90 genes varies greatly across species. For instance, while the two human cytoplasmic Hsp90 genes have only low-frequency non-synonymous mutations (Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA), the four Arabidopsis thaliana cytoplasmic Hsp90 genes exhibit a much higher degree of polymorphism (Ossowski et al. 2008). Knock-outs of any of the *A. thaliana* cytosolic Hsp90s have few phenotypes, suggesting redundancy among the gene copies and providing an explanation for the higher degree of

polymorphism. In *S. cerevisiae* there are two cytosolic Hsp90s. Most strains had identical cytosolic Hsp90s, but for both forms, some strains had truncating deletions of one of the two copies (Engel et al. 2013). Slight variations in function may exist in the different copies of Hsp90 or across polymorphic individuals, which is consistent with balancing selection maintaining variation in Hsp90 function. Though Hsp90 interacts with many co-chaperones that regulate its function and specificity, the amount of variation present in Hsp90 suggests the gene itself also has experienced pressure to diversify functionally.

Hsp90 is a buffer of genetic variation

Hsp90 is a hub of biological networks because of its biochemical function as a protein chaperone of a large set of clients. Not surprisingly, perturbation of Hsp90 function affects a large number of qualitative and quantitative phenotypes across organisms. In *Drosophila melanogaster* (Rutherford & Lindquist 1998), *A. thaliana* (Queitsch et al. 2002), *Danio rerio* (Yeyati et al. 2007), and *S. cerevisiae* (Jarosz & Lindquist 2010) a large number of pleiotropic effects were observed when Hsp90 function was impaired by pharmacological inhibition, moderate increase in temperature, or genetic mutation. These altered phenotypes were present at low frequency, but the diversity of phenotypes was broad. Remarkably, the phenotypes were specific to the genetic background, and functional Hsp90 was described as buffering this genetic variation from reaching phenotypes.

Rutherford & Lindquist (1998) found further support that the phenotypes observed were due to differences in the genetic backgrounds of the flies. The trait revealed by inhibition of Hsp90 could be inherited across generations in multiple independent lines, indicating that the

effect was unlikely to be due to stochastic errors in development. Multigenerational selection for individuals possessing the trait resulted in independent populations with variable penetrance of the trait. Lines with high penetrance were outcrossed with unselected lines, and the trait penetrance in the resulting offspring was greatly reduced. Chromosomes harboring the genetic variation underlying these differences were identified, indicating that the genetic variation responsible for increased penetrance was present prior to selection and revealed by perturbation of Hsp90.

Similar results were observed in *A. thaliana*, which has a vastly different levels and patterns of genetic variation (Queitsch et al. 2002). *A. thaliana* is an inbreeding plant with low levels of heterozygosity, especially when compared to *D. melanogaster* (Leffler et al. 2012). Recombinant inbred lines (RIL), which differ genetically between lines, but are nearly homozygous within a line, were derived from two accession of *A. thaliana*, and examined for their phenotypes upon Hsp90 inhibition. Repeatedly, the same traits were observed within a line when Hsp90 was inhibited, but diverse traits were observed across lines. As plants are sessile organisms, their responses to changing environments are exquisitely fine-tuned. Examination of the RIL responses to different conditions demonstrated that Hsp90 buffers genetic variation in plasticity as well (Queitsch et al. 2002).

The next step in understanding Hsp90's role in buffering genetic variation was to identify the variants that are buffered. Gibson and Dworkin (2004) suggested a number of approaches to identify buffered loci, including quantitative trait loci (QTL) mapping. In this strategy, a conditional quantitative phenotype is measured in RILs in both in the causal condition and in a control condition. The response of the trait to the causal condition is QTL mapped, identifying

loci that are responsible for the response (T Sangster et al. 2008). In the case of Hsp90-buffered loci, the causal condition would be one in which Hsp90 is perturbed. To determine the frequency of Hsp90-buffered variation, *A. thaliana* control RILs and RILs in which Hsp90 function was reduced through pharmacological inhibition or small interfering RNA were measured for 23 traits (T Sangster et al. 2008; Todd A Sangster et al. 2008). Across traits, on average 0.5 QTL were mapped per chromosome per trait on three of the five *A. thaliana* chromosomes, demonstrating that Hsp90 buffered-loci are common in the genome and common to phenotypes (T Sangster et al. 2008).

Specific genetic variants that interact with Hsp90 were identified in *S. cerevisiae* where mapping is easier. One-hundred traits were measured across 104 RILs in both the presence of an Hsp90 inhibitor and control conditions. Traits that mapped to a locus in the control but not the Hsp90-inhibited condition were said to be buffered, whereas loci that were only mapped in the Hsp90-inhibited condition were said to be potentiated. Regardless of the terminology, Hsp90-interacting loci were mapped for nearly every trait (Jarosz & Lindquist 2010), further supporting the findings in *A. thaliana* that Hsp90 interacting variation is common (T Sangster et al. 2008). The underlying variants for four of the QTL were identified. Three of the four were found in the coding regions of proteins; of those three proteins, two were known to interact with Hsp90. The fourth locus mapped to a variant in a 3' untranslated region (UTR). This variant was associated with much higher RNA levels at this locus when Hsp90 was inhibited, suggesting that Hsp90 interacts with a protein that regulates gene expression through contact with the 3'UTR (Jarosz & Lindquist 2010). Apparently the genetic variation buffered by Hsp90 does not have to be

harbored by genes that encode Hsp90 clients, making the potential for Hsp90 to influence the genome vast.

The buffering of cryptic genetic variation

Genetic variation buffered by Hsp90 is a form of cryptic genetic variation: standing genetic variation that influences phenotypes conditionally (Paaby & Rockman 2014). Cryptic genetic variation does not influence phenotypes in all conditions, which complicates efforts to map the genetic variation underlying disease and may partially explain the phenomenon of "missing heritability" (Queitsch et al. 2012; Gibson 2012). Missing heritability refers to the failure of the cumulative heritability of identified causal variants to sum to the observed heritability of a trait. Similarly, cryptic genetic variation may underlie threshold traits, only having effects in the presence of certain genetic variants (Hartl & Clark 2007).

Both genetic and environmental perturbations can reveal cryptic genetic variation. Cryptic genetic variation is revealed through genetic perturbation because of the interactions that exist between genes. Genetic interactions can be classified into two types: one gene-by-one gene (epistasis) or one gene-by-many genes (buffering). An example of epistasis would be when a common allele in one gene that singly has no effect on phenotypes can alter phenotypic expression once combined with another common allele (Gibson & Dworkin 2004). Hsp90's ability to interact with many other genes is an example of buffering. There are a number of mechanisms by which these genetic interactions can be perturbed. The genetic background can change due to sexual reproduction, exposing the phenotypic effects of cryptic genetic variation. Alternatively, novel mutations may expose cryptic genetic variants. Furthermore, a change in the

environment can expose the phenotypic effects of cryptic genetic variation. For example, variation in the body size of the three spine stickleback increases upon transfer from seawater to freshwater through a revelation of cryptic genetic variation (McGuigan et al. 2011).

Several properties of genetic networks may define which genes harbor cryptic genetic variation (Wagner 2005). Genes that are redundant with others may harbor cryptic genetic variation. For example, if two genes share function, genetic variation may not reach the phenotype they both regulate. However, if both genes harbor variants, phenotypes may be affected. In cases of redundant genes, it may be difficult to discern which gene harbors the cryptic genetic variation. It may be that both do. Genes involved in feedback control also may act similarly to harbor genetic variation. If a gene that regulates the level of an output is affected, but there is feedback control as well, variation in the regulatory genes may be cryptic. However, if both the regulatory and the feedback control genes are affects, the cryptic genetic variation may be exposed. This concept is also referred to as "regulatory buffering" (Masel & Siegal 2009). Buffering has been suggested to be an emergent property of complex networks that cannot be traced to a particular mechanism (Siegal & Bergman 2002), suggesting that any gene can harbor cryptic genetic variation. In the case of Hsp90 and the mechanism by which it buffers genetic variation, it seems more likely that genes encoding clients would harbor cryptic genetic variation.

On the other hand, which genes in genetic networks buffer cryptic genetic variation? Hubs in genetic networks have been suggested to reveal cryptic genetic variation because they interact with a larger number of other genes (Rutherford & Lindquist 1998; Lehner et al. 2006; Levy & Siegal 2008), although evidence also suggests that all genes in complex networks can

expose cryptic genetic variation (Siegal & Bergman 2002). These opposing hypotheses can be resolved if one considers degree. The search for genes that allow for the storage and release of cryptic genetic variation has the goal to find genes that can buffer genetic variation broadly, in many other genes. These genes would be classified as strong genetic modifiers—buffering many other genes (Paaby & Rockman 2014). At the same time, other genes can be in epistasis with only one or few other genes, yet still buffer cryptic genetic variation (Milloz et al. 2008). Genes that have broad biochemical roles, such as regulators of chromatin, miRNA, splicing, or even DNA damage control are expected to buffer genetic variation broadly (Lempe et al. 2012). However, as found by Bergman and Siegal (2003), all genes may have the potential to buffer cryptic genetic variation because of the pervasiveness of epistasis. In fact, the situation even can be thought of in terms of the potential epistasis between nucleotide base-pairs. The first cryptic genetic variant mapped was found in the gene 3' end of the gene *Egfr* in *D. melanogaster* interacting with a variant upstream in *Egfr*, regulating the photoreceptors of the eye (Dworkin et al. 2003). Even a locus within the same gene can buffer cryptic genetic variation.

Evolutionary effects of chaperone-buffered cryptic genetic variation

It has been proposed that Hsp90 is an evolutionary capacitor, storing genetic variation and exposing it to selection when Hsp90 function is perturbed. As Hsp90's function is regulated by environmental conditions, potentially adaptive novel phenotypes will be revealed to selection simultaneous to environmental change (Rutherford & Lindquist 1998). Notably, cryptic genetic variation may be of a higher quality than novel mutations; modeling suggests that cryptic genetic variation, if not perfectly neutral, may be exposed to "pre-selection" and the most deleterious cryptic variants purged. These variants are said to be pre-adaptive compared to novel mutations

(Masel 2006). As Hsp90 acts as a chaperone, stabilizing its clients, we may expect that cryptic genetic variation tolerated by clients may be more deleterious than variation tolerated by non-clients.

Several pieces of evidence suggest that Hsp90-dependent genetic variation has been revealed throughout evolution. First, an increased correlation between genotype and phenotype is observed when Hsp90 function is perturbed (Jarosz & Lindquist 2010), supporting that Hsp90 increases the relative importance of genetics over environment. Second, one naturally occurring adaptation has been tied to Hsp90's role in buffering genetic variation. In cavefish, loss of the eye is adaptive. The precursor population to cavefish inhabits surface water and has functioning eyes. When Hsp90 in these surface fish is inhibited, an increase in the variation in eye size is observed, presumably due to release of cryptic genetic variation. The authors found that decreased conductivity of water—an environmental trigger—resulted in an increased variance in eye size in surface dwelling fish similar to that observed upon Hsp90 inhibition (Rohner et al. 2013). Thus, an environment regulating Hsp90's buffering of cryptic genetic variation was identified.

A specific evolutionary signal of Hsp90 buffering genetic variation has been detected in the genes that encode Hsp90 clients. Kinase clients of Hsp90 have a greater dN than kinase non-clients, indicating that Hsp90 client accumulate more non-synonymous changes over evolutionary time compared to non-clients (Taipale et al 2012). Notably Hsp90 is not the only chaperone for which such enhanced evolutionary rates in substrates have been observed.

Chaperones may influence the evolutionary rates of their substrates by stabilizing their clients (Taipale et al. 2012, 2010). Increased protein stability is positively correlated with the evolvability of proteins (Bloom et al. 2006). The increased stability provided by chaperones may allow for a greater tolerance of mutations, as most mutations destabilize proteins (DePristo et al. 2005). Based on this hypothesis, mutations that affect amino acids in chaperoned proteins would be increased relative to mutations affecting synonymous sites; however this increase in non-synonymous changes would not be observed in genes encoding non-clients.

The bacterial chaperone, GroEL/ES has been implicated in promoting the evolution of its substrates. Overexpression of GroEL/ES increased the variance in enzymatic activity in mutagenized copies of proteins. Destabilizing mutations and mutations to the protein core were tolerated due to the overexpression of GroEL/ES, which allowed for greater range in enzyme function (Tokuriki & Tawfik 2009). In several studies, a correlation between evolutionary rate and GroEL/ES dependence has been sought (Williams & Fares 2010; Bogumil & Dagan 2010). Importantly, many other factors correlate with the evolutionary rate of proteins as measured by the ratio of non-synonymous changes to synonymous changes, dN/dS. Genes that are highly expressed have lower dN/dS compared to genes that are less expressed (Pál et al. 2001; Wall et al. 2005). The position of proteins in protein networks is also associated with evolutionary rate. Proteins with many interactions have decreased dN/dS compared to protein with few interactions (Fraser et al. 2002). After correcting for such factors, a correlation between GroEL/ES dependence and evolutionary rate of substrates was observed. Although Taipale et al (2012) completed a similar analysis for Hsp90 and its substrates, no corrections were made for these potentially confounding factors. It has been observed that clients tend to have lower protein

abundance, so it possible that the difference in dN between clients and non-client kinases can be explained through expression level. Although examining dN or dN/dS does not address Hsp90's ability to buffer non-coding variation, it does provide evidence as to whether Hsp90 has acted as an evolutionary capacitor.

Hsp90 as a buffer of developmental noise

Not only does Hsp90 buffer genetic variation broadly, but also it dampens developmental noise from phenotypes. In the former role, Hsp90 allows genetic variation to accumulate with an environmentally-dependent mechanism of release that can facilitate evolution. In the latter role, functional Hsp90 reduces stochastic errors of development, likely through its contribution to maintaining cellular homeostasis. Whether this second role also facilitates evolution is less clear.

The measure of developmental noise of a phenotype is known as robustness and is a quantitative trait. Traditionally, robustness of individuals has been measured as the degree of symmetry in morphological features (Clarke 1998). Another robustness measure is the degree of accuracy with which a genotype produces a phenotype across many isogenic siblings, as described by the variance of the trait. Robustness thus measured is trait-specific and may not be predictive of robustness in other traits (Clarke 1998). Like any quantitative trait, robustness shows a distribution among genetically divergent individuals of a species and can be mapped to distinct genetic loci (Hall et al. 2007; Jimenez-Gomez et al. 2011; Todd A Sangster et al. 2008).

Robustness is commonly attributed to features of underlying genetic networks, such as connectivity, redundancy, feedback, and oscillators (Whitacre 2012; Casanueva et al. 2012; Masel & Siegal 2009). Targeted perturbation of these features decreases phenotypic robustness.

One of the best characterized 'master regulators of robustness' is the molecular chaperone Hsp90 (Casanueva et al. 2012; Todd A Sangster et al. 2008; Burga et al. 2011; Gangaraju et al. 2011; Jarosz & Lindquist 2010; Queitsch et al. 2002; T Sangster et al. 2008; Sollars et al. 2003; Yeyati et al. 2007; Hsieh et al. 2013; Specchia et al. 2010). Hsp90 inhibition decreases robustness to developmental noise in plants and yeast (Todd A Sangster et al. 2008; T Sangster et al. 2008; Rutherford & Lindquist 1998; Queitsch et al. 2002; Jarosz & Lindquist 2010; Yeyati et al. 2007; Hsieh et al. 2013) (Figure 1.2 a, b, Figure 1.3 a, b). Hsp90's capacity to buffer noise in many developmental phenotypes has been attributed to its high connectivity in genetic networks (Sangster et al. 2004).

Perturbing Hsp90 function impairs its numerous substrates, which is thought to alter network connectivity and lead to decreased robustness to developmental noise. Indeed, pharmacological or small RNA inhibition of Hsp90 or Hsp90 mutation in isogenic *A. thaliana* results in an increase in variance in a quantitative trait and reveals a variety of novel phenotypes (Queitsch et al. 2002; Samakovli et al. 2007). When Hsp90 is inhibited, robustness as measured by symmetry is also affected in zebrafish. Yeyati et al (2007) observed asymmetries in the eyes of zebrafish when Hsp90 was perturbed. Results in flies were conflicted: in some traits, symmetry was affected by inhibition of Hsp90 (Debat et al. 2006; Milton et al. 2006) whereas other traits were not (Milton et al. 2003).

Mutational penetrance is a common theme in relation to Hsp90 function. In isogenic populations grown in the same environment, mutational penetrance refers to the variability in which a mutation is expressed in the phenotype due to differences in developmental noise. Queitsch et al (2002) observed rare phenotypes in RILs that were more frequent when Hsp90

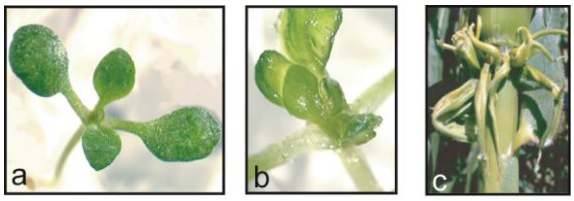

Figure 1.2. Loss of robustness in developmental traits. a) A wild-type *A. thaliana* seedling shows normal phyllotaxy and leaf shape. b) An HSP9-reduced *A. thaliana* seedling shows distorted phyllotaxy and meristem defects. c) A *Zea mays* RNA polymerase IV mutant shows abnormal lateral outgrowths.

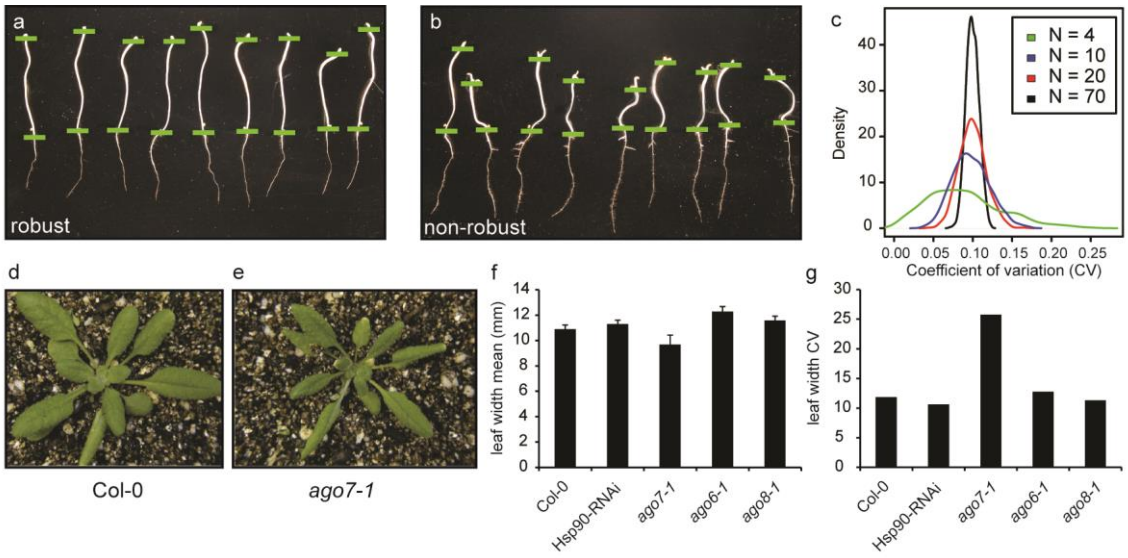

Figure 1.3. Using quantitative traits to measure robustness. a-b) The early seedling trait hypocotyl length is a sensitive robustness read-out. a) robust 7 day-old, dark-grown wild-type *A. thaliana* seedlings, b) non-robust 7 day-old, dark-grown mutant *A. thaliana* seedlings. c) Coefficient of variation (CV, standard deviation/mean) is an unreliable estimate of population robustness at small sample sizes. Here, we show the distribution of CV estimates for various sample sizes from simulated data. For each sample size N, 1000 samples of size N were drawn from a normal distribution with mean = 10 and SD = 1 (true CV = 0.1), and a CV estimate was calculated for each sample. d-e) *ago7* mutants show decreased robustness in leaf width. Representative plants of d) wild-type Col-0 and e) *ago7-1*. f) Mean width and g) CV of the longest leaf at flowering time was measured in Col-0, an HSP90-reduced line (HSP90-RNAi), and *ago6-1*, *ago7-1*, and *ago8-1*, n=15.

was inhibited. Similarly, Yeyati et al (2007) found that inhibition of Hsp90 affected the penetrance of some mutations but not all. In *C. elegans*, naturally varying protein levels of Hsp90 has been associated with mutational penetrance. Levels of GFP-fused Hsp90 protein were assayed using microscopy in individual worms that had a variably penetrant mutation. Lower levels of Hsp90 protein was measured in worms that expressed the mutant phenotype compared to those with a wild-type phenotype (Burga et al. 2011). Hsp90 buffers both quantitative and qualitative traits from developmental noise. Bet-hedging is one strategy taken by organisms to deal with changing environments. As naturally varying levels of Hsp90 function affects developmental noise, Hsp90 ability to buffer noise may be an example of a bet-hedging strategy.

Hsp90 is not the only known robustness master regulator. In yeast, a systematic mutant analysis examined hundreds of microscopy-based phenotypes to generate a global measure of robustness for each mutant. Three-hundred robustness master regulators (Levy & Siegal 2008) were identified. A correlation was found between the level of phenotypic robustness provided by a gene and the number of proteins interacting with gene products, indicating that these regulators tend to be hubs in genetic networks. Chromatin organizers and transcription regulators were enriched in this set of robustness master regulators. Interestingly in this study, the protein chaperone Hsp70 was identified as a master regulator of robustness, whereas the two Hsp90 homologs in yeast were not.

The circadian regulator *ELF4* is another gene that reduces robustness when perturbed (Doyle et al. 2002). Circadian clocks are endogenous oscillators with remarkably robust periods, which persist in the absence of light cues and under increased temperature (Mas & Yanovsky 2009). The robustness of plant clocks is thought to arise from multiple interconnected feedback

loops (Mas & Yanovsky 2009). In reporter assays, *elf4* mutants show highly variable periods before turning arrhythmic (without periods) (Doyle et al. 2002). It is unclear whether the initial, variable periods translate into increased variation of developmental traits, but it seems likely given the importance of the circadian clock in orchestrating growth and development. In fact, Hsp90's effect on robustness may arise in part from disrupted clock function: ZTL, a circadian regulator, is chaperoned by Hsp90 (Kim et al. 2011).

Fine-tuning of gene expression stabilizes developmental traits.

The presence of feedback loops within networks promotes robustness (Masel & Siegal 2009). In 2006, Hornstein and Shomron hypothesized that microRNAs (miRNAs) may reduce gene expression noise and sharpen developmental transitions. In particular, feed-forward loops, in which a transcription factor regulates both a target and its miRNA with opposing effects on target protein levels, were predicted to buffer stochastic expression fluctuations (Hornstein & Shomron 2006). miRNAs have recently been shown to facilitate robustness. For example, *miRNA164* miRNAs control plant development by dampening transcript accumulation of their targets *CUC1* and *CUC2*, wherever expression of miRNAs and targets overlap. *miRNA164* miRNAs define boundaries for target mRNA accumulation in addition to reducing target expression levels (Sieber et al. 2007).

In plants, small RNA-dependent regulation of gene expression is not limited to miRNAs – in fact, there are many plant-specific small interfering RNAs (siRNAs), some of which are mobile and facilitate robust pattern formation. Chitwood and co-authors (2009) demonstrated that a subset of trans-acting siRNAs (tasiRNAs), the low-abundant and conserved tasiR-ARFs,

move intercellularly from the upper leaf side (adaxial), where they originate, to the lower leaf side (abaxial), generating a small RNA gradient that defines the expression boundaries of the abaxial determinant ARF3. tasiR-ARF biogenesis requires both miRNA activity (*miR390*) and siRNA pathway components, including the specialized Argonaute AGO7. Although *miR390* accumulates in a seemingly non-specific pattern throughout the developing leaf, tasiR-ARF biogenesis is restricted to the most adaxial leaf cell layers by the localized expression of *AGO7* (Chitwood et al. 2009). Consistent with the notion that the tasiR-ARF gradient mediates robust adaxial-abaxial fate decision, *ago7* mutants show significantly increased variance in adaxial leaf width (Figure 1.3 d-g).

Systematic searches for new master regulators of robustness

Hall and co-authors (Hall et al. 2007) mapped the first quantitative trait loci (QTL) for trait robustness rather than trait mean in two recombinant inbred populations (RILs), estimating within-genotype robustness with Levene's statistic. They identified 22 robustness QTL across five developmental traits in two conditions. Of these, only three QTL affected exclusively trait robustness, whereas all the others coincided with mean QTL. This strong correlation of robustness and mean QTL agrees with Waddington's view that decreased robustness is associated with decreased function (Hall et al. 2007). Nearly half of the robustness QTL were linked to *ERECTA*, for which a mutant allele segregated in both RILs. *ERECTA* controls aerial organogenesis, and *erecta* mutants show strong pleiotropic phenotypes (Shpak et al. 2004, 2005). Using the same approach, Sangster and colleagues (Todd A Sangster et al. 2008) mapped robustness QTL for two seedling traits, including a third RIL population without a segregating *erecta*-allele. Most robustness QTL did not coincide with mean QTL under control conditions.

As in the previous study, heritability of robustness QTL was significantly lower than for mean QTL. Both studies provided empirical evidence for network elements that stabilize particular traits, which was subsequently also shown in maize (Ordas et al. 2008).

Moving from developmental traits to large-scale molecular traits, Jimenez-Gomez and colleagues (2011) mapped robustness QTL in Bay x Sha RILs for defense metabolite levels and genome-wide expression. For both datasets, the authors were limited to about four replicates per line for estimating within-genotype robustness with the coefficient of trait variation (CV, standard deviation/mean). CV of small samples is an unreliable robustness estimate (Figure 1.3 c). As CV is strongly mean-driven, the identified robustness QTL may be largely driven by mean differences. Countering this concern, the authors point out that not all mean QTL coincided with robustness QTL. However, all but one robustness QTL coincided with mean QTL. The authors addressed the sample size problem by mapping line-specific CV averages for all 22,746 transcripts, identifying loci that affect global gene expression CV. The major effect QTL contained *ELF3*, an important circadian and flowering time regulator. The *ELF3* alleles differentially affected robustness for some traits but not others, possibly due to buffering of the significant global CV expression differences (Jimenez-Gomez et al. 2011). The *ELF3* alleles produce significant mean differences in circadian and developmental phenotypes in reference backgrounds (Coluccio et al. 2011; Jimenez-Gomez et al. 2011). In contrast to *ELF4* (Doyle et al. 2002), *elf3* loss-of-function mutants do not show decreased robustness in circadian or developmental phenotypes (Coluccio et al. 2011; Jimenez-Gomez et al. 2011), although the ELF4 and ELF3 proteins are known to interact (Nusinow et al. 2011).

Together, these studies prove that robustness is a quantitative trait with strong genetic contributions. In reference or hybrid backgrounds, natural alleles can cause different robustness levels in specific traits. In their natural backgrounds, however, these low robustness alleles may be buffered by compensatory mutations. Most *A. thaliana* robustness QTL were trait-specific, suggesting that natural robustness alleles reside in trait-specific sub-networks rather than in robustness master regulators that destabilize many traits and reduce fitness when perturbed. For example, an *Hsp90* allele that subtly decreases robustness in some traits has been found in wild flies, yet this slightly deleterious allele is exceedingly rare (Sgrò et al. 2010; Chen & Wagner 2012).

Most recently, Shen and colleagues (2012) developed a statistical framework to associate trait variation across accessions with genetic variation (vGWAS), re-analyzing a published *A. thaliana* GWA dataset. Their analysis found loci in which allelic variation is associated with accessions that either vary little from each other in trait mean or vary greatly from each other (Fig. 1.4). The authors demonstrate that accounting for these variance or penetrance differences significantly improves heritability and hence mapping of trait means. By eliminating invariant accessions and focusing on those with variable penetrance one could identify background-specific trait determinants (Shen et al. 2012). The successes of QTL mapping for robustness regulators and vGWAS suggest that GWAS for novel robustness regulators will be successful.

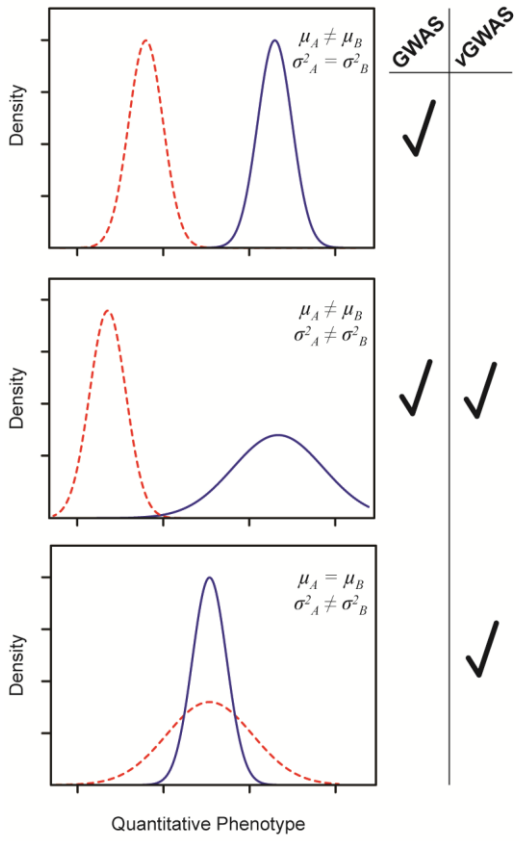

Figure 1.4. vGWAS loci that cannot be mapped with GWAS. Traditional GWAS associates significant mean differences in quantitative traits with genetic variants (top and middle panel). In contrast, vGWAS associates significant variance differences in trait means with common genetic variants in the same individuals. This approach does not consider within-genotype robustness, rather it is a measure of penetrance of a given polymorphism. Alleles associated with higher variance may or may not affect the trait mean (middle and bottom panels respectively). Thus, taking variance effects into account may improve heritability estimates relative to traditional GWAS.

Evolutionary theory of phenotypic robustness

Theory predicts that fluctuating environments may encourage the selection for greater variance in traits while selection in stable environments results in lower trait variance (Masel & Siegal 2009). To model selection on variance controlling genes, Pettersson and colleagues (2013) created a simple system in which phenotypes of individuals were determined from the input of a variance-controlling gene, several effector loci and the rest of the genome. The variance-controlling locus had two alleles: a high variance allele and a low variance allele. Directional selection acted first and more strongly on the variance-controlling gene than on the effector loci which only affected the mean. As intuition suggests, directional selection resulted in the fixation of the high variance allele in the simulations. Additionally, higher rates of environmental change resulted in selection for the high variance allele of the variance-controlling gene (Pettersson et al. 2012), consistent with the reduction in variance observed in stabilizing selection experiments (SanCristobal-Gaudy et al. 2001). This theoretical results suggest that regulators of developmental noise are important to the genetic architecture underlying phenotypes.

Focus of the dissertation

How phenotypes are buffered from genetic and stochastic perturbations and when this buffering is evolutionarily relevant are poorly understood. Can the genetic buffering of Hsp90 be detected in measures of evolutionary rate? Can such a signal be detected across species? These questions will be addressed in Chapters 2-4. Hubs like Hsp90 provide robustness to phenotypes from developmental noise. Is the robustness provided by these hubs emergent from networks or can it be explained by a particular genetic interactor? Do non-hub genes buffer developmental

noise? These questions will be addressed in Chapter 5. In Chapter 6, I commence a study to identify novel regulators of developmental noise. In total, these studies add to our knowledge about Hsp90's role as a buffer to both genetic and stochastic developmental perturbations.

CHAPTER II

THE PROTEIN CHAPERONE HSP90 CAN FACILITATE THE DIVERGENCE OF GENE DUPLICATES[1]

Abstract

Heat shock protein 90 (HSP90) acts as a chaperone by ensuring the proper maturation and folding of its client proteins. The HSP90 capacitor hypothesis holds that interactions with HSP90 allow proteins to accumulate mutations while maintaining function. Following this logic, HSP90 clients would be predicted to show relaxed selection compared with non-clients. In this study, we identify a new HSP90 client in the plant steroid hormone pathway: the transcription factor BES1. Its closest paralog, BZR1, is not an HSP90 client. This difference in HSP90 client status in two highly similar proteins enabled a direct test of the capacitor hypothesis. We find that *BES1* shows relaxed selection compared to *BZR1*, hallmarks of neo- and subfunctionalization, and dynamic HSP90 client status across independent evolutionary paths. These results suggested that HSP90's influence on gene evolution may be detectable if we compare gene duplicates, because duplicates share most other properties influencing evolutionary rate that might otherwise conceal the chaperone's effect. We test this hypothesis using systematically-identified HSP90 clients in yeast, and observe a significant trend of HSP90

[1] This chapter is published as "The protein chaperone Hsp90 can facilitate the divergence of gene duplicates," Genetics: 193(4), 2013, pp. 1269-1277, by J Lachowiec, T Lemus, JH Thomas, PJJ Murphy, JL Nemhauser, C Queitsch.

clients evolving faster than their non-client paralogs. This trend was not detected when yeast clients and non-clients were compared without considering paralog status. Our data provide evidence that HSP90 influences selection on genes encoding its clients and facilitates divergence between gene duplicates.

Introduction

The phenotypic capacitor HSP90 is thought to influence evolutionary processes through its ability to both conceal and release genetic variation (Jarosz & Lindquist 2010; Queitsch et al. 2002; Rutherford & Lindquist 1998; Yeyati et al. 2007). Perturbation of this conserved and essential chaperone reveals cryptic genetic and epigenetic variation in flies, plants, fish, and yeast (Jarosz & Lindquist 2010; Queitsch et al. 2002; Rutherford & Lindquist 1998; Yeyati et al. 2007). In worms, HSP90 affects the penetrance of partial loss of function mutations (Burga et al. 2011). As expected under the capacitor hypothesis, worms with naturally lower HSP90 levels show significantly higher mutation penetrance (Casanueva et al. 2012). HSP90-dependent variation can be revealed by moderate environmental stress alone, providing a plausible release mechanism for this concealed variation in nature (Jarosz & Lindquist 2010; Queitsch et al. 2002; Rutherford & Lindquist 1998). We showed previously that HSP90-dependent variation is common in natural plant populations, implicating the chaperone as an important player in shaping phenotype and evolutionary trajectories (Todd A Sangster et al. 2008; T Sangster et al. 2008). Together, these findings have prompted a longstanding debate about the importance of HSP90 in evolutionary processes and the magnitude of its effect (Bergman & Siegal 2003; Meiklejohn & Hartl 2002; Rando & Verstrepen 2007).

It is well-established that HSP90 recognizes metastable proteins and facilitates their folding and stability (Taipale et al. 2010). Recent studies demonstrate that protein stability is a major constraint on protein evolution (Bloom et al. 2006; Pena et al. 2010). Stable proteins tend to evolve faster as they can explore greater sequence space without losing function (Bloom et al. 2006). In prokaryotes, overexpression of the chaperonin GroEL/ES allows the evolution of a far greater number of highly active enzyme variants by compensating for their reduced stability (Tokuriki & Tawfik 2009). We hypothesized that in eukaryotes HSP90 may facilitate gene divergence by similarly relaxing constraints on protein stability. If so, HSP90 clients should show greater tolerance to mutations and evolve faster than non-clients facing similar evolutionary pressures. To address HSP90's role in gene evolution in this context, we focused on recent gene duplicates, which initially face similar selection pressures and encode proteins of similar stability. We used the brassinosteroid (BR) pathway in *Arabidopsis thaliana* as an experimental model because every step in BR signaling is encoded by gene families (Kim & Wang 2010) and previous studies suggested that the BR pathway may require HSP90, although no HSP90 client had been identified (Sangster et al. 2007; Sangster & Queitsch 2005).

We demonstrate here that only one of two paralogous transcription factors in the BR pathway is an HSP90 client and that its encoding gene shows relaxed purifying selection compared to its non-client paralog. Gene duplicates diverge through sub- and neofunctionalization. Consistent with subfunctionalization, only the HSP90 client is temperature sensitive; consistent with neofunctionalization, the gene encoding the HSP90 client contains a novel exon and non-synonymous polymorphisms in divergent *A. thaliana* strains. HSP90-

facilitated divergence of gene duplicates is widespread, because in the yeast *Saccharomyces cerevisiae*, genes encoding HSP90 clients tended to evolve faster than those encoding their non-client paralogs. Together, our data provide strong evidence for HSP90-faciliated evolution in extant genomes and hence strong support for the capacitor hypothesis.

Methods Summary

Plant growth conditions and treatments. Columbia-0 (Col-0) was used as wild type (WT). *bes1-D*, *bzr1-D*, *DWF4*-ox, *bin2-1*, *BRI1*-ox, and *bes1-2* (WiscDsLox246D02) were in the Columbia-0 background. Seedlings were grown for seven days on media with DMSO (mock) or geldanamycin, brassinolide, and brassinazole, dissolved in DMSO. Statistical significance of response of hypocotyl length of seedlings from 2-3 replicates of 10-60 seedlings was determined using standard least square linear regression.

Biochemistry. For western blot, seven-day-old seedlings, grown in red LED light, were ground in liquid nitrogen. Buffer (0.15M Tris pH 6.8) was added, and extracted protein was quantified using Bradford's assay. Proteins were resolved using SDS-PAGE, transferred to nitrocellulose and probed with anti-BES1 antibody. For co-immunoprecipitation, ground rosette tissue was used. Extracted protein was incubated with Protein L Agarose, which was pre-incubated with anti-HSP90 3G3 antibody. Beads were pelleted and washed in buffer. Anti-BES1 antibody was used to detect BES1 in the input and pellet.

Phylogenetic tree and dN/dS analysis. Sequences for *BZR/BEH* family members in available sequenced plants were acquired from http://phytozome.net v5.0 from a BLAST search for gene families with similarity to *BES1*. MUSCLE 3.7 was used for amino acid alignment of

the identified sequences, and Gblocks was used to remove regions with poor conservation. The remaining 86 sequences were re-aligned and neighbor-joining was used to create a distance tree. For the *BZR/BEH* tree, the outgroup was identified as a *BZR/BEH* family member that was closely related, but an outgroup to all *A. thaliana BZR/BEH* family members. Sequences were aligned in MUSCLE 3.7 and PhyML was used for maximum likelihood tree (Guindon & Gascuel 2003). For dN/dS analysis, codeml from PAMLv4.4b was run using models 0, 1, and 2 (Bielawski & Yang 2003).

Yeast data analyses. Published HSP90 interactors were used (Zhao et al. 2005). The branch length of HSP90 interactors in three-member and two-member families was obtained from Ensembl Compara (release 61).

Results

BES1 is an Hsp90 candidate

Inhibition of HSP90 yields a wide variety of morphological phenotypes in *A. thaliana* plants (McLellan et al. 2007; Queitsch et al. 2002; Sangster et al. 2007; Whitesell et al. 1994). Among these phenotypes, we previously noted severely dwarfed plants, which closely resembled known BR mutants. To directly test whether the BR pathway requires HSP90 function, we grew seedlings in the presence of exogenous BR (brassinolide, the most biologically active BR) with and without the highly specific HSP90 inhibitor geldanamycin (GdA) (Queitsch et al. 2002). Inhibition of HSP90 function significantly interfered with response to BRs (Fig. S1A, R^2=0.68, $p<0.0001$, linear regression model, standard least square fit). Consistent with this finding, GdA

also reduced seedling response to brassinazole, an inhibitor of BR biosynthesis (Asami et al. 2000), (Fig. S1B, red light, R^2=0.76, p<0.0001; fig. S1C, dark R^2=0.81, p<0.0001).

We next addressed what step in the BR signaling pathway was most responsive to a loss of HSP90 function. The best-characterized HSP90 clients are the mammalian steroid hormone receptors and kinases (Picard et al. 1990; Taipale et al. 2010; Whitesell et al. 1994). The most common clients are transcription factors (Taipale et al. 2010). In *A. thaliana*, only a few clients are known, none of which function in the BR pathway (Hubert et al. 2003; Iki et al. 2010; Ishiguro et al. 2002; Takahashi et al. 2003). As HSP90 clients do not share a common sequence or structural motif, client status is typically determined by a combination of genetic and biochemical analyses. Here, we took advantage of several well-characterized mutants in the BR pathway to test their response to HSP90 inhibition. We focused on the most likely clients: the steroid hormone receptor kinase BR INSENSITIVE1 (BRI1) and the transcription factors BES1 and BZR1 (Wang et al. 2001). To distinguish between HSP90 effects on BR signaling versus BR synthesis, we included a mutant in DWARF4 (DWF4), an enzyme that catalyzes a rate-limiting step of BR biosynthesis. Well-characterized gain-of-function mutants were used to bypass the extensive redundancy in the BR pathway. As had been shown previously each mutant significantly increased hypocotyl length, (He et al. 2005), (Fig. 1A, B). Upon inhibition of HSP90 with 0.5 µM GdA, *BRI1-ox*, *DWF4-ox*, and *bzr1-D* seedlings responded like wild-type seedlings (*BRI1-ox*, p=0.22; *DWF4-ox*, p=0.99, *bzr1-D*, p=0.6273), whereas *bes1-D* seedlings showed significant hypersensitivity to HSP90 inhibition (p<0.0001, table S1, fig. 1A-C). Notably, the dominant *bes1-D* and *bzr1-D* mutants carry the identical amino acid change from proline to leucine (Tang et al. 2011; Wang et al. 2002; Yin et al. 2002), yet their response to

HSP90 inhibition differed dramatically. Further increases in GdA levels yielded a significant response in *BRI1-ox*, *DWF4-ox* (*BRI1-ox,* $p<0.0001$; *DWF4-ox,* $p=0.0003$), but not in *bzr1-D* seedlings ($p=0.5918$, table S1), suggesting that there may be additional HSP90 targets upstream of the transcription factors (Fig. 1 A, B). Similar results were obtained in the dark (Fig. S2A, B). Together these data suggest that BES1, but not BZR1, is an HSP90 client.

Another criterion for HSP90 client status is proof of physical interaction (Taipale et al. 2010). Using a co-immunoprecipitation assay with an HSP90-specific antibody, we found that BES1 physically interacts with HSP90 in plants (Fig. 1D). In the absence of BRs, BES1 is negatively regulated by BIN2 and several related kinases (Vert & Chory 2006). Cellular perception of BRs triggers inhibition of BIN2 and activates BES1. Hypophosphorylated, active BES1 can be detected as a fast mobility band on western blots (Yin et al. 2002) (Fig. 1E). In this form, BES1 interacts with other transcription factors, binds DNA, and promotes plant growth (Yin et al. 2005). In our studies, BES1 can interact with HSP90 independent of phosphorylation, as we were able to pull-down faster and slower mobility BES1 bands (Fig. 1D).

While BRs and GdA have opposite effects on plant growth, and likely on BES1 function, inhibition of HSP90 produced a similar shift in BES1 mobility as BR treatment (Fig.1E). Like other well-established clients, such as the *Drosophila* Argonaute Piwi and the human transcription factor HSF, BES1 was not degraded upon HSP90 inhibition (Fig. 1E) (Gangaraju et al. 2011; Zou et al. 1998). However, the hypophosphorylated BES1 that accumulated upon HSP90 inhibition was unable to promote growth (Fig. 1A-D). Our data suggest that BES1 may require HSP90 for its activation, perhaps by facilitating BES1 dimerization, promoting nuclear translocation, or interfering with phosphorylation by BIN2. Following this logic, inhibition of

HSP90 should enhance any gain in BIN2 function. *bin2-1* mutants are semi-dominant hypermorphs with repressed BR signaling and strongly reduced BES1 activity (Kim & Wang 2010). As predicted, loss of HSP90 activity sensitized seedlings to a gain of BIN2 activity, as evidenced by a dramatically increased proportion of severely dwarfed seedlings in a segregating population of *bin2-1* mutant seedlings (Fig. S2C). These results also support that BIN2 is a non-client.

HSP90 facilitates the divergence of gene duplicates

If HSP90 indeed allows its clients to explore a wider range of sequence space, the *BES1* gene would be predicted to show evidence of relaxed selection compared to the *BZR1*gene. We created a phylogenetic tree of the *A. thaliana BZR*/*BEH* gene family using a gene from *Aquilegia coerulea* as an outgroup (Fig. 2A, S3). *BES1* and *BZR1* are the most recently diverged paralogs among the six *A. thaliana BZR*/*BEH* family members, with 88% amino acid identity (Wang et al. 2002). We determined the ratio of the rate of non-synonymous substitutions to the rate of synonymous substitutions (dN/dS) for *BES1* and *BZR1*, under a model allowing all branches of the *A. thaliana BZR*/*BEH* tree to evolve at different rates. Consistent with our prediction, the gene encoding the HSP90 client BES1 (dN/dS=0.09) shows relaxed purifying selection compared to the gene encoding the non-client BZR1 (dN/dS=0.04) (Fig. 2A). As *BES1* and *BZR1* diverged recently and only differ in a small number of amino acids, their difference in dN/dS was not significant. However, we took advantage of the entire *BEH* gene family to test whether *BES1* shows a different evolutionary rate compared to other family members. Using a maximum likelihood approach (Bielawski & Yang 2003), we showed that *BES1* indeed exhibits a significantly different evolutionary rate, if all other branches are assumed to evolve at the same

rate (2δ=5.232, df=1, p=0.0222). In contrast, no significant difference was found for the evolutionary rate of *BZR1* under the same assumptions (2δ=3.244, df=1, p=0.0717).

To address whether HSP90 clients generally evolve faster, we analyzed a data set of systematically identified *Saccharomyces cerevisiae* HSP90 clients (Zhao et al. 2005). Proteins that physically interact with HSP90 by tandem affinity purification-tagged (TAP) mass spectrometry (Zhao et al. 2005) were considered as likely HSP90 clients. Likely HSP90 co-chaperones, identified by the TPR domain (Wegele et al. 2004), were removed from the analysis. Unlike clients, co-chaperones interact with HSP90 through the TPR domain and modulate HSP90 activity. As expected, the evolutionary rates of HSP90 clients did not differ significantly from all other yeast genes (Fig. S4A). Factors such as differences in selection pressure, protein stability, or codon bias, among others, likely obscure any impact on evolutionary rate by HSP90. To compare genes well-matched for these factors, we identified gene duplicates in which one paralog encoded a likely HSP90 client. Consistent with our results for *BES1*, genes encoding yeast HSP90 clients showed significantly longer branch length than their respective closest paralog (Fig. 2B, Table S2, n=13, 95% confidence interval 1.02-1.64, p=0.002, one-sample Wilcoxon test, testing the deviation from the expected ratio of client/non-client branch length of 1). Likely non-clients, identified by synthetic genetic interaction with an *hsp90* mutation (SGI), did not show this trend (Fig. S4B, n=27, p=0.97). We then tested a more stringent situation. In cases in which yeast HSP90 interaction status was the derived state (i.e., not present in a common ancestor), three out of four HSP90 clients showed longer branch length than their respective closest paralog (Fig. 2B, three-member families). Next, we addressed whether these differences in evolutionary rate were due to expression differences between clients and their respective non-

client paralogs. Genes that evolve faster tend to be expressed at a lower level (Drummond et al. 2005). In contrast, many clients are significantly higher expressed than their non-client paralogs across nearly 200 environmental conditions (Gasch et al. 2000) (Table S2, Fig. S5). We observed no correlation of expression levels and branch lengths between clients and their respective paralogs (Fig. S4C, Fig. S5, Table S2). Taken together, our data suggest that HSP90 can facilitate the divergence of gene duplicates in yeast and plants.

HSP90 client BES1 shows hallmarks of sub-and neofunctionalization

Evolutionary theory holds that after gene duplication, one copy dies off quickly or changes function (Conant & Wolfe 2008). A surviving gene copy can retain part of the ancestral gene function, such as expression in fewer tissues or under certain environmental conditions (subfunctionalization) and/or acquire a novel beneficial function (neofunctionalization) (Conant & Wolfe 2008). An obvious subfunctionalization path for an HSP90 client is loss of function under environmental conditions that challenge HSP90 chaperone activity, such as increased temperature. HSP90 clients are typically less stable than other proteins and hence lose function at increased temperature despite induced HSP90 expression (Taipale et al. 2010). We grew seedlings at 27ºC, a temperature known to challenge HSP90 function but not induce heat stress in *A. thaliana* (Queitsch et al. 2002). The temperature response of *BRI1-ox*, *DWF4-ox*, and *bzr1-D* mutants closely resembled the response of wild-type seedlings (Fig. 3D). In contrast, not one of more than 100 *bes1-D* seeds in multiple independent experiments germinated at 27ºC (Fig. 3D, E). This germination phenotype was completely suppressed at standard growth conditions (22ºC) (Fig. 3D, E, Fig. S2D). The loss of BES1 function at moderately elevated temperature, likely a direct result of challenged HSP90 function, is strong support for subfunctionalization.

If genes encoding HSP90 clients evolve faster than their non-client paralogs, neofunctionalization may be facilitated. Although there appears to be extensive redundancy between the HSP90 client BES1 and the non-client BZR1 (Yin et al. 2005), we found evidence of BES1 neofunctionalization. First, the major *BES1* splice variant At1g19350.3 encodes a novel exon not found in *BZR1* or any of the other *BEH/BZR* family genes (Fig. 3A). This exon shares significant homology with intergenic regions on chromosomes 3 and 5 (Fig. 3B). This exon together with adjacent intron sequence has additional matches to the 5'UTR *BZR1* sequence and to another intergenic region (Fig. 3C). Gene chimeras are a hallmark of neofunctionalization (Hahn 2009), and these findings suggest that distant genomic regions may have contributed to the novel *BES1* exon.

Second, we found that *BES1* polymorphisms across divergent *A. thaliana* accessions were significantly associated with the phenotypic variation these strains showed in response to HSP90 inhibition. Wild *A. thaliana* accessions harbor considerable genetic variation, yet due to *A. thaliana's* inbreeding life-style, individual accessions are nearly isogenic. Sensitivity to HSP90 inhibition varied dramatically among accessions (Fig. 4A). Some accessions grouped with the hypersensitive *bes1-D* mutant (Fig. 4A, red bar), whereas others responded very little. We found significant associations between sensitivity to HSP90 inhibition and two *BES1* polymorphisms (Fig. 4B, S6A). The first polymorphism is a frameshift mutation in the strains Zdr-6 and Tottarp-2 that results in an early stop codon (Fig. 4A, black bars, fig. 1E). Consistent with a severely truncated BES1 protein, these strains responded little to HSP90 inhibition. The second polymorphism was found in four different strains (Fig. 4A, green bars). This synonymous polymorphism in the 5' end of *BES1* alters the preferred codon for alanine to a rarely used codon

(Wright et al. 2004). Changes in codon usage can alter translation efficiency and protein folding kinetics, potentially making stabilization by HSP90 superfluous. A similar change in codon usage has been observed for evolutions of viral proteins under conditions of reduced HSP90 in mammalian cells (Vaughan et al. 2010). Thus, in strong support of the capacitor hypothesis, HSP90 inhibition revealed phenotypic differences among divergent *A. thaliana* strains that are associated with *BES1* polymorphisms. Moreover, HSP90 client status appears to be highly dynamic with BES1 losing HSP90 dependence in some strains.

Another non-synonymous polymorphism in *BES1* did not correlate with response to HSP90 inhibition (Fig. 4B). Consistent with HSP90 facilitating *BES1* evolution, this polymorphism may represent a step towards a novel phenotype through acquisition of a second mutation, with which it interacts epistatically (Fig. 4B) (Salverda et al. 2011). Neutral non-synonymous mutations, such as mutations that increase protein stability without affecting function, can facilitate ascent to new fitness optimum (Tokuriki & Tawfik 2009). In contrast, all twenty *BZR1* polymorphisms were synonymous (Fig. 4B), consistent with *BZR1*'s lower evolutionary rate compared to *BES1*. None of these were associated with sensitivity to HSP90 inhibition, supporting BZR1's non-client status (Fig. 4B, S6B). Eight *BZR1* polymorphisms were unique to the strain Uod-7, which carries a large intronic insertion in addition to other intronic polymorphisms (Fig. 4B). These polymorphisms may lead to mis-regulation of *BZR1* (Le Hir et al. 2003), thereby increasing the need for BES1. This scenario is consistent with Uod-7's hypersensitivity to GdA (Fig. 4A). Taken together; our data suggest that HSP90 facilitates divergence of gene duplicates by promoting sub-functionalization through temperature sensitivity of its clients and neofunctionalization through their increased tolerance of mutations.

Discussion

With BES1, we have identified a novel HSP90 client in a crucial plant growth pathway (Fig 1). By necessity, plant growth must be finely tuned to the environment. The temperature sensitivity of BES1 tightly links BR pathway function to the ambient environment. We and others showed previously that HSP90 plays an important role in defenses against herbivores and microbial pathogens, (Hubert et al. 2003; Sangster et al. 2007; Sangster & Queitsch 2005; Takahashi et al. 2003), as well as the timing of flowering (Sangster et al. 2007). Our findings add an important layer of complexity to known hormone-environment interactions (Nemhauser 2008; Robert-Seilaniantz et al. 2011), especially in light of HSP90's known role in resource allocation among defense and growth pathways. Our data also add to the emerging evidence of functional divergence between BES1 and BZR1—the critical downstream targets of BR signaling. For example, recent genome-wide chromatin immunoprecipitation experiments show only partial overlap of gene targets in BES1 and BZR1 (Sun et al. 2010; Yu et al. 2011). This partial overlap highlighted key unresolved questions about BR transcriptional responses. Specifically, what distinguishes BES1- or BZR1-specific targets from targets regulated by both proteins, and what determines whether BES1 and BZR1 act as repressors or activators? As BES1, but not BZR1, is an HSP90 client, we speculate that interaction with the chaperone may facilitate association of BES1 with specific partner proteins, resulting in BES1-specific functions (Yin et al. 2005).

In addition to identifying a novel plant HSP90 client, our study provides support for the capacitor hypothesis by garnering evidence of HSP90-facilitated evolution in extant genomes. When we compared paralogs, we found a significant trend that genes encoding diverse HSP90 clients evolved faster than non-clients (Fig. 2). In contrast, when we compared diverse yeast

HSP90 clients and non-clients as aggregate groups, we could not detect any significant difference in evolution rate (Fig. S4B). This result is consistent with a similar analysis comparing prokaryotic genes encoding diverse clients of the bacterial chaperonin GroEL/ES and non-clients (Williams & Fares 2010). On the contrary, within the superfamily of mammalian kinases, strong HSP90 clients carry more non-synonymous mutations than non-clients (Taipale et al. 2012). These different results are not surprising because many other factors or gene properties influence evolutionary rate, most of which will differ for a diverse set of genes. Combined, these factors can outweigh and conceal effects of HSP90 and GroEL/ES client status on evolutionary rate. By focusing on gene duplicates of diverse HSP90 clients, which share many properties influencing evolutionary rate, we found support for HSP90's previously hypothesized effect on gene evolution. As overexpression of GroEL/ES also increases the evolutionary rates of its clients (Tokuriki & Tawfik 2009), our study supports an ancient and conserved role for protein chaperones in gene evolution.

Our findings also help resolve an apparent paradox about the fate of gene duplicates. Gene duplicates are functionally redundant immediately after duplication, rendering one copy superfluous or even harmful (Lynch & Conery 2000; Papp et al. 2003). Evolutionary theory predicts non-functionalization—one gene copy is silenced and subsequently lost—as the fate of most duplicated genes. Recent studies have challenged this view by revealing that gene duplicates with partially redundant function are maintained much longer than expected (Conant & Wolfe 2008; Dean et al. 2008; DeLuna et al. 2008; Lynch & Conery 2000; Maere et al. 2005). Moreover, in yeast, plants, insects, and humans, genes that can be maintained in duplicate show strong functional bias, with transcription factors and kinases significantly over-represented (Aury

et al. 2006; Conant & Wolfe 2008; Guan et al. 2007; Maere et al. 2005; Wapinski et al. 2007). Both observations raise questions about the molecular mechanism(s) that aid the initial preservation, continued maintenance, and eventual divergence of a specific subset of gene duplicates. Our data are consistent with a model in which acquisition of HSP90 client status is a molecular mechanism aiding each of these steps. Acquisition of HSP90 client status can occur through a single mutational step (Citri et al. 2006; Taipale et al. 2010). A new HSP90 client will be subject to immediate environmental subfunctionalization. As we observed for BES1, HSP90 clients are exquisitely sensitive to environmental conditions affecting protein folding (Nathan et al. 1997). Immediate and efficient subfunctionalization will counteract deleterious dosage effects and foster long-term maintenance of gene duplicates, and hence provide opportunity for their functional divergence. In fact, *bes1-D* mutants are more sensitive to an increase in temperature than to HSP90 inhibition with GdA. This hypersensitivity also suggests that multiple protein chaperones are necessary for BES1activity. In addition to providing evidence for HSP90-faciliated subfunctionalization, we show that HSP90 client status correlates with hallmarks of neofunctionalization and increased evolutionary rates of the genes encoding them. HSP90 recognizes metastable signal transduction proteins, most of which are transcription factors and kinases (Taipale et al. 2010). We speculate that HSP90's specificity for these substrates contributes to the observed functional bias among gene duplicates. Additional support for HSP90's role in gene duplicate divergence comes from our observation that HSP90 client status is dynamic in the BZR/BEH family and in yeast gene families, with frequent gains and losses (Fig. 4A, B). Dynamic client status is also observed in the mammalian kinome (Taipale et al.

2012). In fact, even across wild *A. thaliana* accessions BES1 HSP90 client status itself appears to be dynamic.

The presence of HSP90 clients and non-clients in gene families with partially redundant function has important implications for another aspect of HSP90-mediated capacitance. Inhibition of HSP90 increases phenotypic variation even in the absence of genetic variation (Queitsch et al. 2002; Todd A Sangster et al. 2008; T Sangster et al. 2008). This increase of phenotypic variation in isogenic lines has been attributed to an increased frequency of stochastic events in development or a greater sensitivity to microenvironments. As recently shown, significant phenotypic variation arises in isogenic worms due to the loss of functional redundancy between paralogs (Burga et al. 2011; Casanueva et al. 2012). Under conditions that challenge HSP90 function, HSP90-dependent paralogs will become inactive, thereby reducing functional redundancy and increasing phenotypic variation. We suggest a general role for HSP90 in maintaining a reservoir of phenotypic variation through facilitating conditional functionality of gene duplicates.

Acknowledgements

This work was supported by grants from the National Human Genome Research Institute Interdisciplinary Training in Genomic Sciences (T32 HG00035) (JL), National Science Foundation Graduation Research Fellowship (DGE-0718124) (JL), National Institute of General Medical Sciences Academic Research Enhancement Award R15-GM086822 (PJMM), and National Science Foundation Grant IOS-0919021 (JLN).

JL, PJMM, JLN, and CQ conceived and designed experiments. JL performed experiments. JL, TL, JHT, and CQ conceived and designed computational analyses. JL and TL performed computational analyses. JL, JLN, and CQ wrote the paper. JLN and CQ should be considered co-corresponding authors on this paper.

We thank members of the Queitsch and Nemhauser groups for useful discussions, Orrin Stone for experimental assistance, Yanhai Yin for anti-BES1 antibody, Elhanan Borenstein, Maitreya Dunham, Stanley Fields, and Harmit Malik for helpful comments, Magnus Nordborg and Scott Hodges for use of unpublished *Aquilegia* sequences, and Neeraj Salathia for manual curation of TPR domain containing proteins.

Main Figures

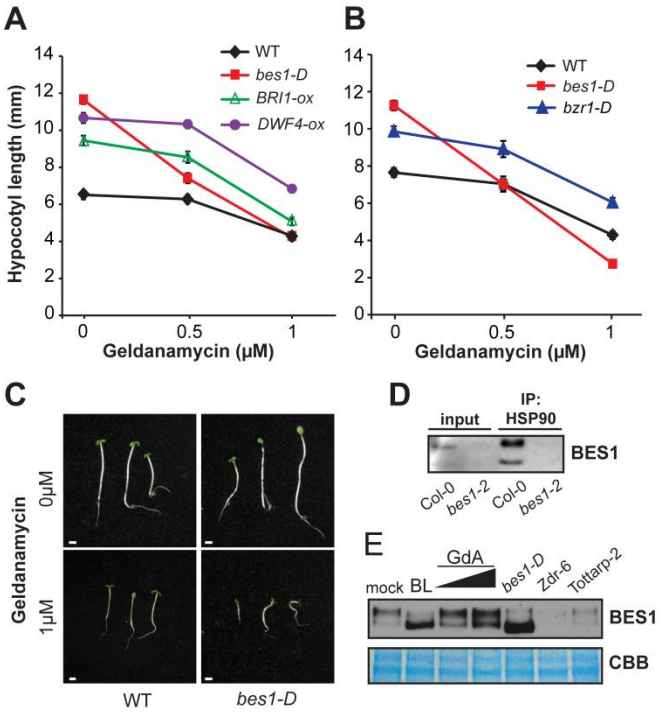

Fig. 1. BES1 is an HSP90 client. (A) Seedlings with increased BR signaling through overexpression of DWF4 (*DWF4-ox*) or overexpression of BRI1 (*BRI1-ox*) showed similar sensitivity to HSP90-inhibition by GdA compared to WT. In contrast, seedlings with constitutive activation of BES1 (*bes1-D*) were significantly more sensitive than WT in red light. Standard error is shown. **(B)** In contrast to the dramatic GdA hypersensitivity of *bes1-D* mutants, *bzr1-D* mutants respond like WT in red light. **(C)** Representative WT and *bes1-D* seedling phenotypes in red light. **(D)** BES1 interacts physically with HSP90. BES1 was immunoprecipitated (IP) with an HSP90 antibody in WT, but not in loss-of-function *bes1-2* mutants, demonstrating the BES1 specificity of the antibody (Swarbreck et al. 2008). Input and IP are shown for both. **(E)** GdA treatment caused a shift of BES1 mobility. Unlike a similar shift caused by brassinolide (BL) treatment, the GdA-induced mobility shift was associated with decreased hypocotyl length. Zdr-6 and Tottarp-2 show reduced levels of BES1; detected protein is presumably due to the presence of other splice forms. Coomassie Brilliant Blue (CBB) is shown as a loading control.

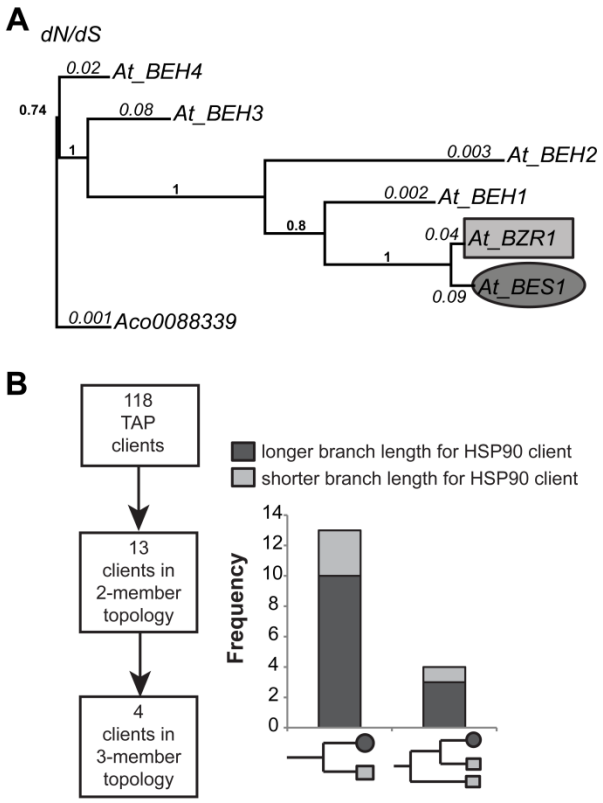

Fig. 2. HSP90 clients show relaxed selection compared to their paralogs. (A) The branch leading to *BES1* has a larger dN/dS ratio than the branch leading to *BZR1*. dN/dS ratios are in italics. Branch lengths represent the difference in the number of amino acid substitutions among family members from an unrooted tree. The fraction of 100 bootstraps supporting each branch are shown in bold. **(B)** *S. cerevisiae* HSP90 clients tend to evolve faster than their paralogs in two-member and three-member families. Clients are dark-grey circles; non-clients are light-grey squares. Star denotes significance. Significance was determined by using a one-sample Wilcoxon test, which tests the deviation from the expected ratio of client/non-client branch length of 1, n=13, 95% confidence interval 1.02-1.64, p=0.0002.

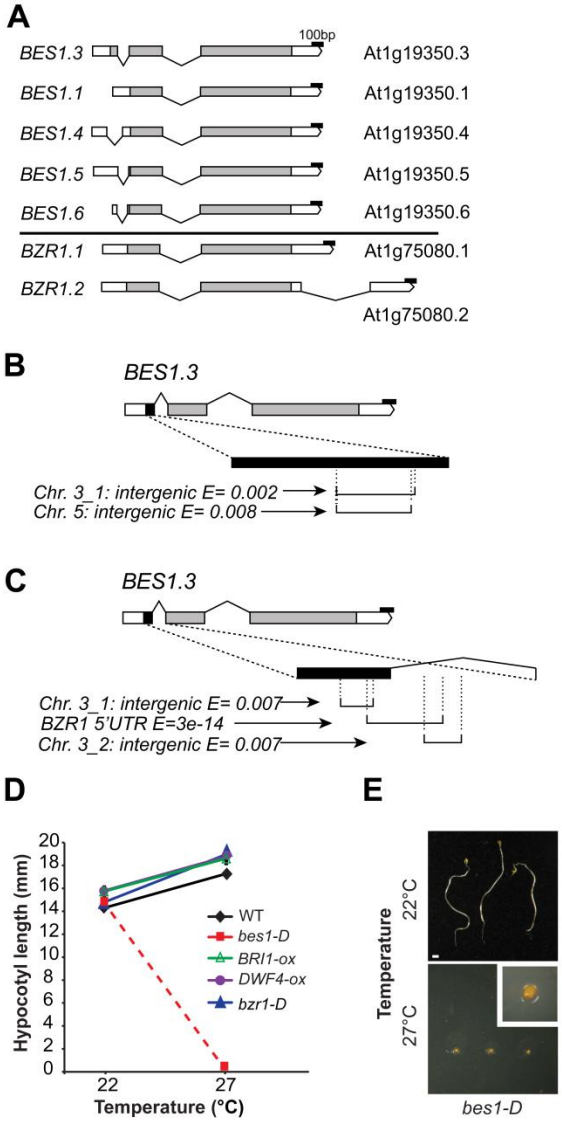
A
100bp
BES1.3
At1g19350.3
BES1.1
At1g19350.1
BES1.4
At1g19350.4
BES1.5
At1g19350.5
BES1.6
At1g19350.6
BZR1.1
At1g75080.1
BZR1.2
At1g75080.2
B
BES1.3
Chr. 3_1: intergenic E= 0.002
Chr. 5: intergenic E= 0.008
C
BES1.3
Chr. 3_1: intergenic E= 0.007
BZR1 5'UTR E=3e-14
Chr. 3_2: intergenic E= 0.007
D
Hypocotyl length (mm)
20
18
16
14
12
10
8
6
4
2
0
22
27
Temperature (°C)
WT
bes1-D
BRI1-ox
DWF4-ox
bzr1-D
E
Temperature
22°C
27°C
bes1-D

Fig. 3. BES1 shows evidence of neo- and subfunctionalization. (A) The major *BES1* splice variant, *BES1.3*, encodes a novel exon. Grey boxes are exons, black lines are introns, and white boxes are UTRs. **(B)** A BLASTn search for regions with homology to the novel *BES1.3* exon identifies intergenic loci. **(C)** A BLASTn search for regions with homology to the novel exon and first intron of *BES1.3* identifies *BZR1* 5' UTR sequence in addition to intergenic loci.**(D)** At 27°C, the *bes1-D* mutant failed to germinate, while all other mutants showed a wild-type response. **(E)** Representative *bes1-D* seedlings and seeds at 22°C and 27°C, respectively. Enlarged image of *bes1-D* (inset) shows a seed that failed to germinate at 27°C.

Fig. 4

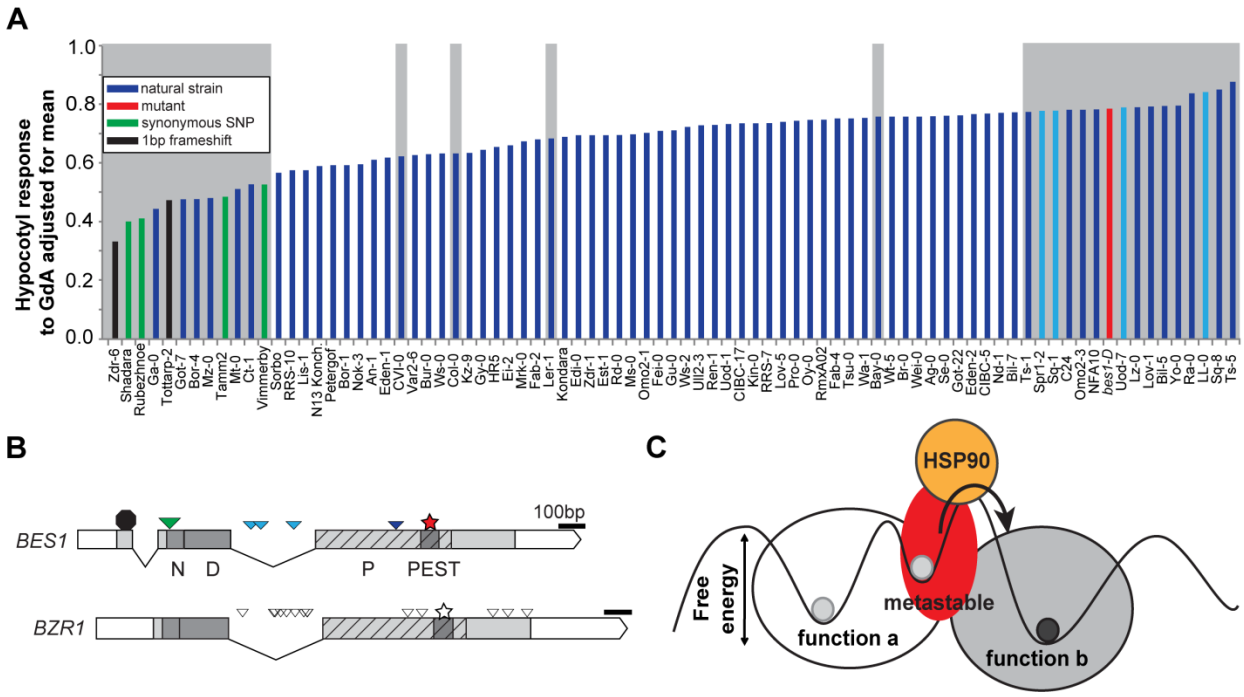
A
Hypocotyl response to GdA adjusted for mean
1.0
0.8
0.6
0.4
0.2
0.0
natural strain
mutant
synonymous SNP
1bp frameshift
B
BES1
100bp
N
D
P
PEST
BZR1
C
HSP90
Free energy
metastable
function a
function b

Fig.4. BES1 HSP90 client status is dynamic.(A) Hypocotyl length GdA sensitivity of divergent *A. thaliana* strains and *bes1-D* (red). *BES1* and *BZR1* were sequenced in strains highlighted in grey. Strains with frameshift mutations are in black; strains with *BES1*-synonymous polymorphisms are green. Both polymorphisms are significantly associated with GdA sensitivity. **(B)** *BES1.3* and *BZR1.1* polymorphisms. Grey boxes are exons, black lines are introns, and white boxes are UTRs. The octagon marks the frameshift mutation, the green triangle marks the synonymous SNP, the blue triangle marks the nonsynonymous polymorphism, white triangles mark non-coding or synonymous polymorphisms, and stars mark the *bzr1-D* and *bes1-D* dominant mutations. The domains are N-nuclear localization signal, D-DNA binding domain, P-phosphorylation domain, PEST-PP2A interaction domain. **(C)** A protein that exists in a free energy minimum (function a) acquires mutations that render it metastable and thus recognized by HSP90. As an HSP90 client, the protein can visit a greater sequence space, increasing the chance of reaching another free energy minimum associated with a novel function (function b) and loss of client status.

Supplementary Materials and Methods

Plant growth conditions and treatments

Columbia-0 (Col-0) was used as wild type. *bzr1-D*, *DWF4*-ox, *bin2-1*, *BRI1*-ox, and *bes1-2* (WiscDsLox246D02) were in the Columbia-0 background. The *bes1-D* mutant was introgressed into the Col-0 background. Previous studies used Columbia wild-type as control (Albrecht et al. 2008; Guo et al. 2009; Li et al. 2009). Seeds were stratified for three nights at 4°C and grown vertically in the dark or red LED light for seven days on 0.5x LS 0.08% bactoagar growth media at 22°C. Geldanamycin (Sigma #G3381), brassinolide (Chemiclones #101), and brassinazole (Chemiclones #117) were dissolved in DMSO. Drugs and mock (DMSO only) treatments were added to growth media. For growth at 27°C, seedlings were placed in 22°C for 24 hours and then transferred to 27°C. For response to GdA in accessions, 15 seeds from 96 accessions were grown in the dark for seven days on medium containing DMSO or 1μM GdA. Accessions in which less than seven seeds germinated were removed from further analysis. Difference between hypocotyl length in mock and treatment was taken and adjusted by the mock mean per genotype.

Statistical analyses

Analyses of data were performed with JMP7 (SAS Institute). To measure the effect of drug treatments and interactions on hypocotyl length, standard least square linear regression was used. Each drug, drug by drug interaction (all fixed effects), and replicate (random effect) was modeled (2-4 replicates for each experiment, 10-60 seedlings each). When comparing the effect of genotype on response to drug treatment, standard least square linear regression was used.

Genotype, drug, genotype by drug interaction (all fixed effects), and replicate (random effect) was modeled (2-4 replicates of each experiment, 10-60 seedlings each). To associate *BES1* and *BZR1* SNPs with response to GdA, the GdA response was normalized using the mock treatment mean for each strain. Each SNP in the 30 strains was tested for association with GdA response using ANOVA (One-way).

Western blot

Seedlings were grown for seven days in red LED light and then ground in liquid nitrogen. Buffer (0.15M Tris pH 6.8) was then added, and extracted protein was quantified using Bradford's assay (Pierce #1856210). Approximately 15ug of protein was loaded per lane. Gel Code Blue Stain Reagent (Pierce #24590) was used to determine loading. Proteins were transferred to nitrocellulose and probed with anti-BES1 antibody (gift of Yanhai Yin).

Co-immunoprecipitation

Col-0 (WT) and *bes1-2* rosette leaves grown in long day conditions were harvested, frozen in liquid nitrogen, ground, and resuspended in HEM buffer (10mM HEPES, 1mM EDTA, 20mM sodium molybdate, 1mM PMSF (Fluka 78830) and protease inhibitor (Roche 11-836-153-001)). The solution was ultracentrifuged for 100,000 x g for 30 minutes at 4°C. The resulting supernatant was incubated for three hours with Protein L Agarose (Pierce 20510), which was pre-incubated with anti-HSP90 3G3 antibody (Enzo Lifescences ALX-804-079-R400). Beads were pelleted and washed with HEM buffer. Anti-BES1 antibody (gift of Yanhai Yin) was used to detect BES1 in the pellet.

Phylogenetic tree and dN/dS analysis

One hundred eighteen sequences for *BZR/BEH* family members in all sequenced plants were acquired from http://phytozome.net v5.0 from a BLAST search for gene families with similarity to *BES1* at the node for Viridiplantae. Sequences from genomes that were not publically available were removed manually. MUSCLE 3.7 was used for amino acid alignment of the identified sequences, and Gblocks was used to remove regions with poor conservation. Sequences with low similarity in the high conservation region identified with Gblocks were removed manually. The remaining 86 sequences were re-aligned and neighbor-joining was used to create a distance tree. For the *BZR/BEH* tree, the outgroup was identified as a *BZR/BEH* family member that was closely related, but an outgroup to all *A. thaliana BZR/BEH* family members. MUSCLE 3.7 was used for amino acid alignment of outgroup and *A. thaliana* family members using default parameters and manually examined for errors. PhyML was used for maximum likelihood tree (Guindon & Gascuel 2003). For dN/dS analysis, PAL2NALv13 was used to convert amino acid alignment to codon alignment. The codeml program from PAMLv4.4b was run with gaps removed using models 0, 1, and 2 (Bielawski & Yang 2003). The branch leading to *BES1* or *BZR1* was allowed to vary in model 2. The dN/dS values for the *BZR/BEH* family came from model 1.

Yeast data analyses

For this study, published HSP90 interactors were used (Zhao et al. 2005). TPR-domain containing proteins were curated from available literature and sequence information and excluded from further analysis. We used the Ensembl Compara (release 61) database to obtain

the branch length of HSP90 interactors in three-member and two-member families. For the three-member families, we filtered out genes in which the inner or outer-paralog were also HSP90 interactors. For the two-member families we filtered out those in which both genes were HSP90 interactors. To determine whether there was a significant difference in branch length between clients, non-clients, and their paralogs, we calculated the proportion of total branch length for each client or non-client and their paralogs. A one-sided Wilcoxon test was used to determine significance of detected differences in proportional branch length between client and paralog or non-client and paralog.

Supplementary figures

Table S1. Significance of response to HSP90 inhibition

	0.5uM GdA		1.0uM GdA	
genotype	R^2	p-value	R^2	p-value
bes1-D	0.73	<0.0001	0.89	<0.0001
bzr1-D	0.26	0.63	0.63	0.59
BRI1-ox	0.48	0.22	0.74	<0.0001
DWF4-ox	0.75	0.99	0.8	0.0003

Table S1. Significance of response to HSP90 inhibition. Seedlings with increased BR signaling through overexpression of DWF4 (*DWF4-ox*) or overexpression of BRI1 (*BRI1-ox*) showed WT-like sensitivity to HSP90-inhibition at 0.5µM GdA. In contrast, seedlings with constitutive activation of BES1 (*bes1-D*) were significantly more sensitive than WT at 0.5µM GdA in red light. At 1.0 µM GdA, *BRI1-ox, DWF4-ox,* but not *bzr1-D* showed increased sensitivity compared to WT.

.

client			non-client paralog			comparison		
gene	expression mean	branch length	gene	expression mean	branch length	client minus paralog expression mean		client branch/non-client branch
YBR172C	-0.3258	1.0072	YPL105C	-0.1654	0.7407	-0.1603	*	1.3598
YCL024W	-0.3181	0.3717	YDR507C	-0.4492	0.2050	0.1311		1.8136
YDL025C	0.3978	0.5842	YOR267C	0.2715	0.4900	0.1263		1.1923
YDL199C	0.5594	1.7737	YFL040W	0.0277	1.7386	0.5317	*	1.0202
YDR001C	0.7171	0.2090	YBR001C	0.5592	0.1304	0.1579	***	1.6024
YFL011W	-0.1650	0.2859	YMR011W	-0.4621	0.1549	0.2971	***	1.8461
YFR024C-A	0.0450	0.3320	YHR016C	0.3144	0.2701	-0.2694		1.2291
YGL077C	-0.3462	0.4449	YNR056C	0.0736	1.0123	-0.4198		0.4395
YHR080C	0.3019	0.5174	YDR326C	-0.3168	0.2316	0.6187	***	2.2340
YKR003W	0.0862	0.2907	YHR001W	0.1561	0.2927	-0.0700	***	0.9931
YMR192W	-0.0287	1.6764	YPL249C	0.0550	0.8823	-0.0837		1.9000
YNR011C	-0.1769	1.0256	YKL078W	-1.1110	1.2136	0.9340	***	0.8451
YNR031C	0.0252	0.5748	YCR073C	-0.2379	0.5540	0.2630	***	1.0375

* p<0.05, ** p<0.01, *** p<0.001

Table S2. Yeast HSP90 clients tend to evolve faster than their respective non-client paralogs independent of expression levels. Yeast HSP90 clients tend to evolve faster than their respective non-client paralogs independent of expression levels. Mean expression was calculated from (Gasch et al. 2000). Differences in mean expression were calculated using a Wilcoxon paired sign rank test. Significance values are abbreviated as * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Fig. S1

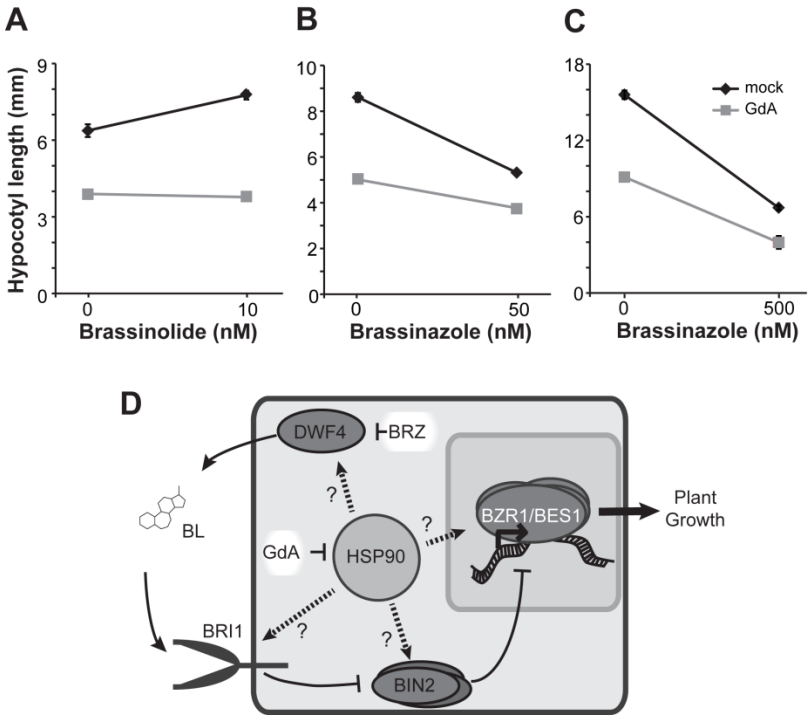
A
B
C
Hypocotyl length (mm)
Brassinolide (nM)
Brassinazole (nM)
Brassinazole (nM)
mock
GdA
D
DWF4
BRZ
BL
GdA
HSP90
BZR1/BES1
Plant
Growth
BRI1
BIN2

Figure S1. HSP90 is required for BR signaling. (A) GdA reduces seedling response to brassinolide (BL), the most biologically active BR in red light. Red light was used because GdA decays rapidly in white light. Standard error is shown for all values. **(B)** Treatment with brassinazole (BRZ), a BR biosynthetic inhibitor, greatly reduces the effect of GdA treatment on seedling growth in red light. **(C)** Treatment with BRZ greatly reduces the effect of GdA treatment on seedling growth in the dark. **(D)** The BR signaling pathway. BRs (brassinolide-BL) are synthesized through a number of enzymatic reactions, including the rate-limiting enzyme DWF4, the target of BRZ. When BR levels are low, BIN2 and related kinases inhibit the activity of a family of transcription factors, including BES1 and BZR1. As BRs accumulate, they are detected by the plasma-membrane localized receptor BRI1. Activated BRI1 triggers a series of phosphorylation and dephosphorylation events that ultimately inhibit the activity of BIN2 and its paralogs. Hypophosphorylated BES1 and BZR1 then can bind DNA and trigger changes in target gene transcription. Each of the indicated proteins is a potential HSP90 client.

Fig. S2

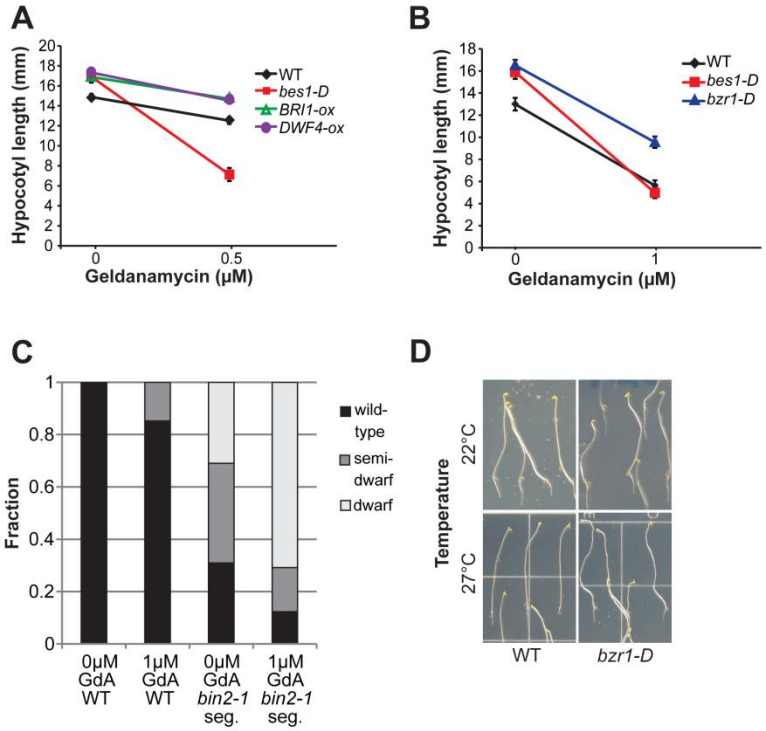
A
Hypocotyl length (mm)
Geldanamycin (μM)
WT
bes1-D
BRI1-ox
DWF4-ox
B
Hypocotyl length (mm)
Geldanamycin (μM)
WT
bes1-D
bzr1-D
C
Fraction
wild-type
semi-dwarf
dwarf
0μM GdA WT
1μM GdA WT
0μM GdA bin2-1 seg.
1μM GdA bin2-1 seg.
D
Temperature
22°C
27°C
WT
bzr1-D

Figure S2. BES1 is an HSP90 client. (A) Seedlings with increased BR signaling through overexpression of DWF4 (*DWF4-ox*) or overexpression of BRI1 (*BRI1-ox*) showed similar sensitivity to HSP90-inhibition by GdA compared to WT in the dark. In contrast, seedlings with constitutive activation of BES1 (*bes1-D*) were significantly more sensitive than WT the dark. Standard error is shown. **(B)** In contrast to the dramatic GdA hypersensitivity of *bes1-D* mutants, *bzr1-D* mutants respond like WT in the dark. **(C)** BIN2 is not a likely HSP90 client. The *bin2-1* mutation results in a dwarfed phenotype compared to WT. The phenotypes of a segregating population of *bin2-1* were enhanced rather than alleviated upon inhibition of HSP90. Plant growth phenotypes were categorized as wild-type, semi-dwarf, or dwarf and assessed for 60 seedlings per genotype and condition. **(D)** Hyper-sensitivity to increased temperature is not observed in WT and *bzr1-D* seedlings grown at 22°C and 27°C.

Fig. S3

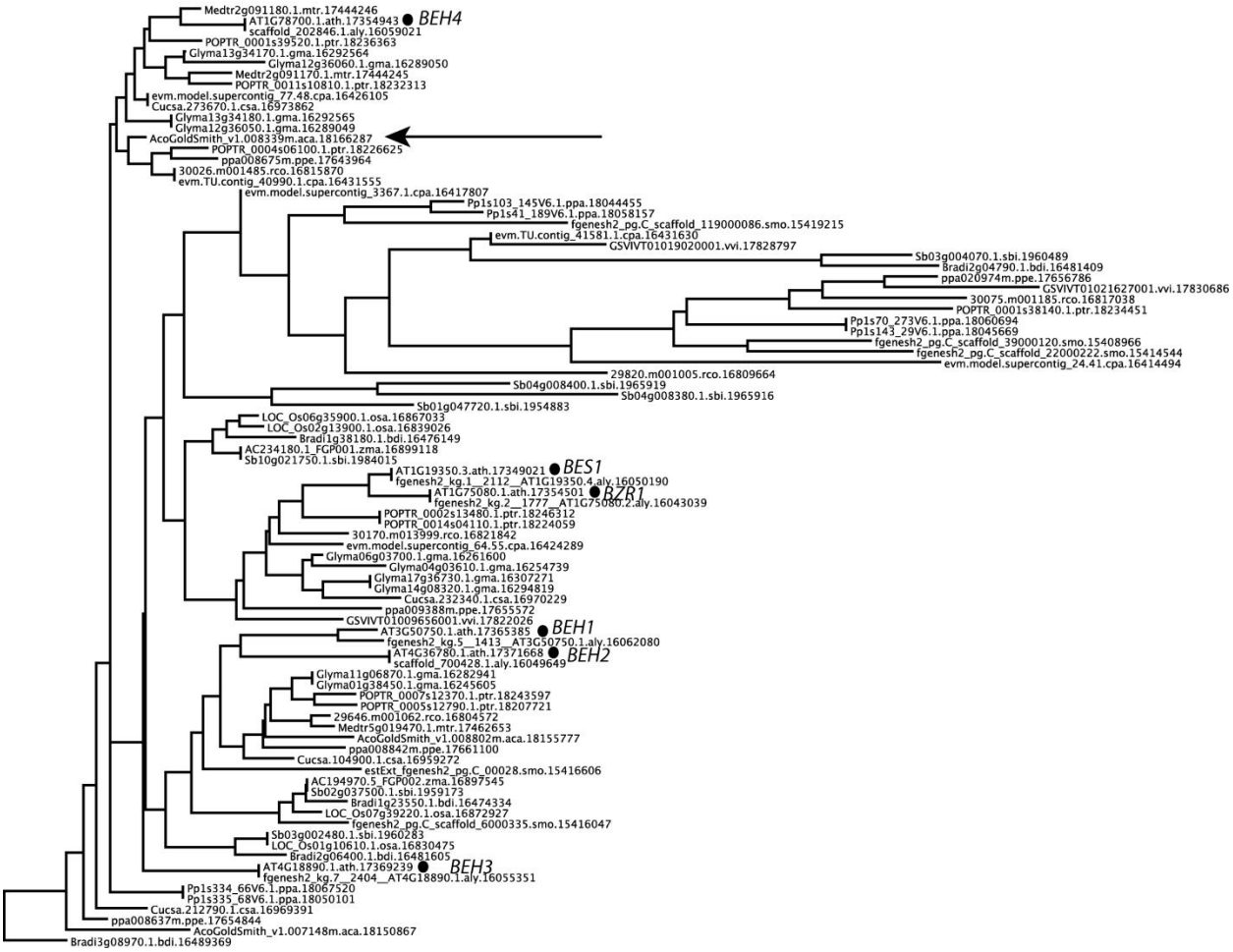

Figure S3. BZR/BEH family genes are found in many plant species. A neighbor-joining tree, using percent amino acid identity among genes identified as potential BZR/BEH family members among 16 plant species, shows that a gene from *Aquilegia coerulea* (arrow) is the most closely related outgroup to all BZR/BEH family members in *A. thaliana* (indicated by "·").

Fig. S4

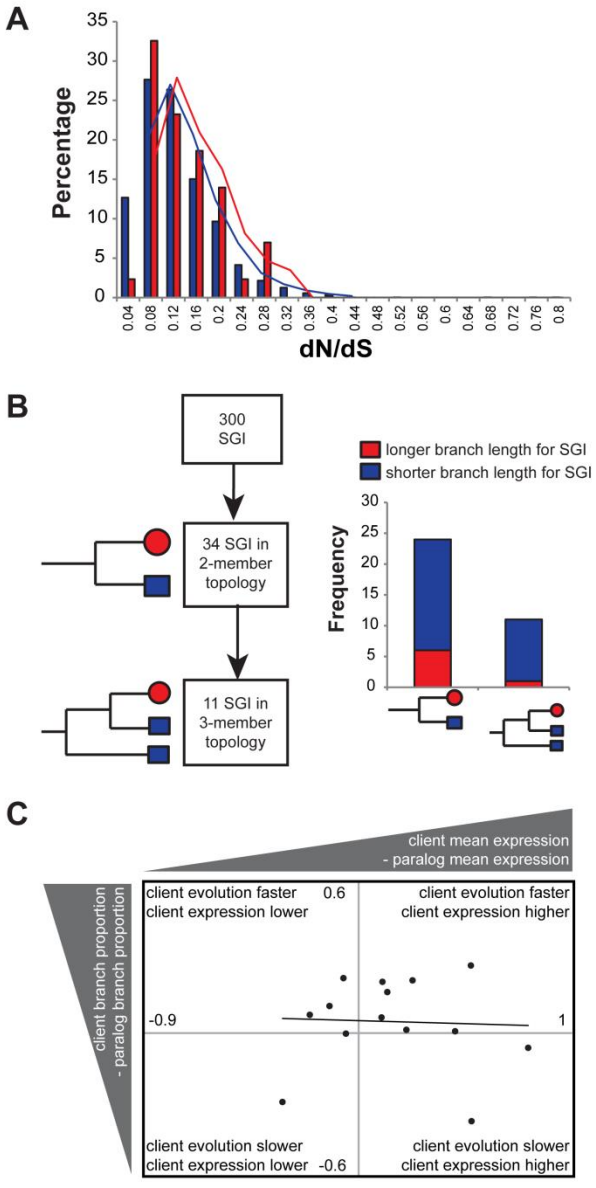
A
Percentage
dN/dS
B
300 SGI
34 SGI in 2-member topology
11 SGI in 3-member topology
longer branch length for SGI
shorter branch length for SGI
Frequency
C
client mean expression - paralog mean expression
client branch proportion - paralog branch proportion
client evolution faster client expression lower
client evolution faster client expression higher
client evolution slower client expression lower
client evolution slower client expression higher
0.6
-0.6
-0.9
1

Figure S4. ***S. cerevisiae*** **HSP90 genetic interactors do not show elevated rates of evolution.** **(A)** *S. cerevisiae* HSP90 TAP-identified clients do not show a significant difference from all other yeast genes in their dN/dS ratios. **(B**) *S. cerevisiae* HSP90 synthetic genetic interactors (SGI) do not show longer branch lengths than their paralogs in two-member and three-member families. In fact, they tend to show significantly shorter branch lengths (p=0.03, Wilcoxon, one-sided). SGI are red circles; SGI paralogs are blue squares. **(C)** Increased evolutionary rate of yeast HSP90 clients is not explained by lower expression. X-axis shows client mean expression minus non-client paralog mean expression across many environmental conditions (Gasch et al. 2000). Y-axis shows client evolutionary rate minus non-client paralog evolutionary rate (using proportional branch length for both). Labels in quadrants indicate relationship of client evolutionary rate and expression level. Numbers indicate scale. Solid trend line shows lack of correlation between evolutionary rate and expression levels (p=0.882, R^2=0.0019).

Fig. S5

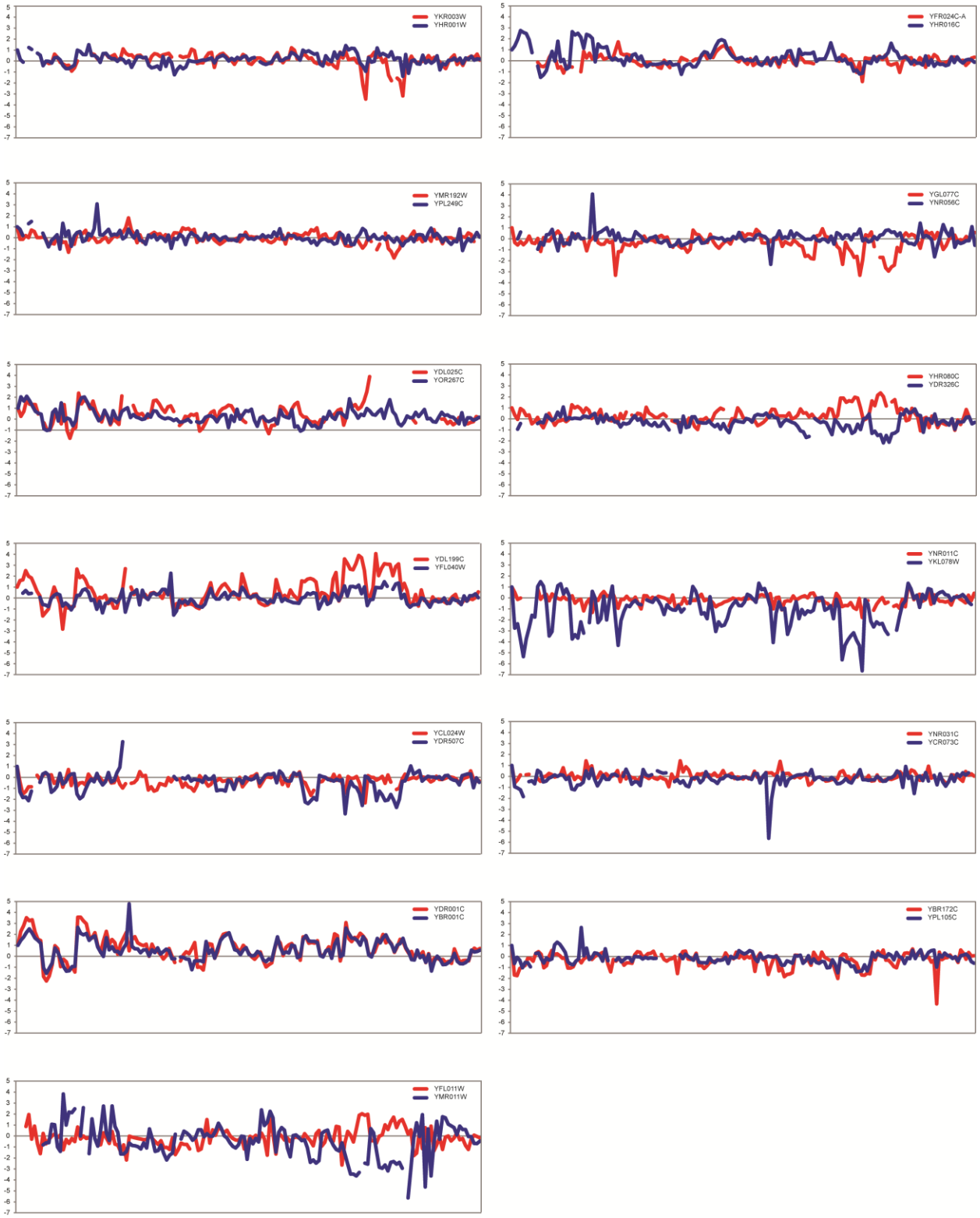
YKR003W
YHR001W
YFR024C-A
YHR016C
YMR192W
YPL249C
YGL077C
YNR056C
YDL025C
YOR267C
YHR080C
YDR326C
YDL199C
YFL040W
YNR011C
YKL078W
YCL024W
YDR507C
YNR031C
YCR073C
YDR001C
YBR001C
YBR172C
YPL105C
YFL011W
YMR011W

Figure S5. Expression profiles of yeast HSP90 clients and their respective non-client paralogs across 173 environmental conditions (Gasch et al. 2000). Client expression is in red; non-client expression is in blue. All 14 pairs are shown. Statistics for observed differences appear in Table S2.

Fig. S6

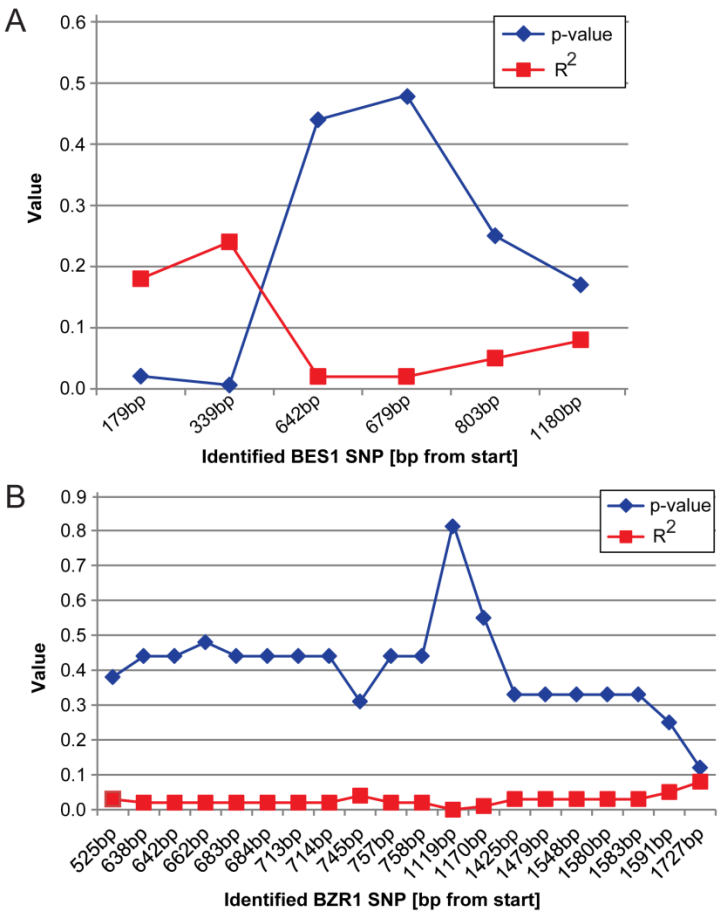

Figure S6. Significant associations were observed between SNPs and hypocotyl response to HSP90 inhibition in *BES1* (A) but not *BZR1* (B) among the 30 sequenced accessions.

ANOVA was used to assess significance.

CHAPTER III

THE PROTEIN CHAPERONE HSP90 AND THE EVOLUTION OF THE HUMAN KINOME[1]

Abstract

Heat-shock protein 90 (Hsp90) promotes the maturation and stability of its client proteins, including many kinases. In doing so, Hsp90 may allow its clients to accumulate mutations as previously proposed by the capacitor hypothesis. If true, Hsp90 clients should show increased evolutionary rate compared to non-clients; however, other factors, such as gene expression and protein connectivity, may confound or obscure the chaperone's putative contribution. Here, we compared the evolutionary rates of many Hsp90 clients and non-clients in the human protein kinase superfamily. We show that Hsp90 client status promotes evolutionary rate independently of, but in a similar magnitude to, gene expression and protein connectivity. Hsp90's effect on kinase evolutionary rate was detected across mammals and increased with time of divergence. Hsp90 clients also showed increased nucleotide diversity and harbored more damaging variation than non-client kinases across humans. These results are consistent with the central argument of the capacitor hypothesis that interaction with the chaperone allows its clients to harbor genetic variation. Hsp90 client status is thought to be highly dynamic with as few as one amino acid change rendering a protein dependent on the chaperone. Contrary to this expectation, we found that across protein kinase phylogeny Hsp90 client status tends to be

[1] This chapter has been submitted to Molecular Biology and Evolution as "Hsp90 promotes kinase evolution" by J Lachowiec, T Lemus, E Borenstein, C Queitsch.

gained, maintained, and shared among closely related kinases. We also infer that the ancestral protein kinase was not an Hsp90 client. Taken together, our results suggest that Hsp90 played an important role in shaping the kinase superfamily.

Introduction

The conserved heat shock protein Hsp90 facilitates the proper folding and stability of its substrates (clients) (Taipale et al. 2010), many of which are kinases with important roles in growth and development. Hsp90 perturbation increases the penetrance of expressed genetic variants and reveals cryptic genetic variation in genetically divergent populations of plant, fly, yeast, and fish (Rutherford & Lindquist 1998; Queitsch et al. 2002; Yeyati et al. 2007; Jarosz & Lindquist 2010). In worms, naturally varying Hsp90 levels predict mutation penetrance with lower Hsp90 levels resulting in greater penetrance (Burga et al. 2011; Casanueva et al. 2012). These observations with traditional model organisms prompted the controversial hypothesis that Hsp90 plays an important evolutionary role, allowing genetic variation to remain phenotypically silent and releasing it in environments that perturb Hsp90 function (Rutherford & Lindquist 1998). Consistent with this hypothesis, perturbing Hsp90 function in surface-dwelling *Astyanax mexicanus* fish results in eye phenotypes that are reminiscent of the natural adaptation of eye loss in the cave-dwelling fish of the same species, presumably due to release of Hsp90-dependent standing variation (Rohner et al. 2013). Hsp90-dependent standing variation occurs frequently in natural strains of plants, flies, and yeast and often affects complex traits (Rutherford & Lindquist 1998; T Sangster et al. 2008; Jarosz & Lindquist 2010), consistent with a significant role of Hsp90 in evolution, especially in the evolution of genes encoding its client proteins.

The evolutionary rate of protein coding genes is commonly measured as the ratio of non-synonymous changes to synonymous changes, dN/dS. Using this measure, we recently reported that genes encoding Hsp90 clients tend to evolve faster than genes encoding their non-client paralogs (Lachowiec et al. 2013). This trend was not observed without considering paralog status, presumably because many other factors, some of which are shared among paralogs, influence evolutionary rate. One drawback of this study was the small number of available client and non-client paralog pairs. Recently Taipale et al. (2012) systematically annotated Hsp90 clients in the human kinome, using a high-throughput assay to assess kinase and Hsp90 interactions. Of the 314 tested kinases, 98 are classified as strong Hsp90 clients and 95 as weak clients. Notably, the dN of strong Hsp90 clients is greater than the dN of non-clients, suggesting that interaction with Hsp90 may indeed allow for increased accumulation of non-synonymous genetic variation (Taipale et al. 2012).

However, several other factors contribute to the evolutionary rate of proteins, including absolute expression levels (Pál et al. 2001; Wall et al. 2005) and a protein's connectivity in protein-protein interaction (PPI) networks (Fraser et al. 2002). Highly expressed genes tend to evolve slower than genes with lower expression levels as do genes encoding proteins that interact with many other proteins, presumably due to the selective pressure to maintain all functional interactions. Neither of the studies discussed above (Taipale et al. 2012; Lachowiec et al. 2013) controlled for these factors when examining the role of Hsp90 on protein evolution. Another factor contributing to dN/dS is protein stability; genes encoding stable proteins tend to evolve faster (Bloom et al. 2006). Interaction with the chaperone Hsp90 promotes client protein stability

as many of its inherently flexible clients degrade upon Hsp90 perturbation (Taipale et al. 2012), suggesting a mechanistic basis for the chaperone's putative effect on dN/dS.

Here we set out to dissect and compare the contributions of Hsp90 client status, gene expression levels and protein interaction degrees to the evolutionary rate of kinases. We evaluate the role of Hsp90 in kinome evolution within and across lineages. Hsp90 client status is thought to be dynamic throughout gene family evolution with as few as one amino acid change resulting in a switch from non-client to client (Citri et al. 2006; Taipale et al. 2012; Lachowiec et al. 2013). We assess this unexplored dynamic by quantifying the transition rates between client and non-client states throughout kinase evolution and show that client status switching is infrequent. We also infer the ancestral state of client status among kinases. Taken together, our results support the controversial hypothesis that Hsp90 plays an important role in the evolutionary processes shaping large gene families.

Results

Hsp90 client status contributes to dN/dS

Taipale and co-authors (2012) previously reported that strong Hsp90 kinase clients acquire more non-synonymous mutations than non-client kinases, using pairwise dN values from human and mouse. We first extended these analyses by examining evolutionary rate (pairwise dN/dS between human and mouse) for non-, weak, and strong clients based on strength of kinase interaction with Hsp90 as defined by (Taipale et al. 2012). Evolutionary rate analysis (dN/dS) is the gold standard for assessing the types of selection potentially acting on proteins. The dN/dS

for strong Hsp90 clients was significantly greater than the dN/dS for non-clients (Figure 1a, $p = 0.0004$, Wilcoxon rank sum test); no significant dN/dS difference was observed between weak clients and non-clients. Both of these findings were consistent with the prior findings observed with dN alone. Second, we found that dN/dS values of non-clients are significantly different from the combined dN/dS values of strong and weak clients ($p = 0.01221$, Wilcoxon rank sum test), which was not previously addressed (Taipale et al. 2012). None of the kinases showed $dN/dS > 1$, which would indicate positive selection. Rather, the Hsp90 client kinases showed relaxed purifying selection compared to non-clients, consistent with previous observations in plants and yeast (Lachowiec et al. 2013).

The observed greater dN/dS of Hsp90 clients may be an indirect consequence of other factors that strongly influence evolutionary rate, specifically gene expression levels (Pál et al. 2001; Wall et al. 2005) or protein interaction degree (Fraser et al. 2002). If genes encoding Hsp90 clients are expressed at lower levels (Taipale et al. 2012) or if client proteins are less connected in protein interaction networks, the correlation between Hsp90 client status and dN/dS can be explained without invoking Hsp90 as a contributor to evolutionary rate. We examined whether gene expression level or PPI connectivity explains the observed contribution of Hsp90 client status to kinase dN/dS, we conducted a linear regression analysis. To this end, we calculated the mean and maximum gene expression of client and non-client kinases across RNA-seq experiments for 11 different human primary tissue samples (Castle et al. 2010) and used PPI values from Taipale et al. (2012). To account for the phylogenetic relatedness among these kinases, we used the kinase tree to calculate phylogenetically independent contrasts (PIC) (Felsenstein 1985) to correct for non-independence of all variables under study: kinase dN/dS,

Hsp90 client status, gene expression levels, and PPI values. For this analysis, Hsp90 client status for each kinase was measured by its quantitative interaction score with the chaperone (Hsp90 interaction score, HIS) (Taipale et al. 2012).

We used linear regression to understand the relationships among the PIC for each pair of variables. The PIC of kinase evolutionary rates and Hsp90 interaction score were positively associated ($p = 0.0001$, Table 1), consistent with our phylogenetically naïve, categorical analysis (Figure 1a). As expected the PIC of kinase evolutionary rate and gene expression levels were negatively associated ($p = 0.02$, Table 1) (Pál et al. 2001); the PIC of kinase evolutionary rate and PPI connectivity were also negatively associated ($p = 0.0001$, Table 1) (Fraser et al. 2002). Notably, we did not observe a significant association between Hsp90 interaction score and expression contrasts ($p = 0.7$, Table 1), suggesting that there are no systematic differences in gene expression levels that may drive the observed differences in evolutionary rate between Hsp90 client and non-client kinases. However, we found a negative correlation between PPI connectivity and Hsp90 interaction score contrasts (Table 1), raising the possibility that the greater evolutionary rate of Hsp90 kinase clients derives from fewer protein interactions.

To further disentangle the contributions of Hsp90 interaction score, gene expression levels, and PPI connectivity to kinase evolutionary rate, we also calculated partial correlations among the contrasts for the four variables. We found that Hsp90 interaction score was positively correlated with kinase evolutionary rate when controlling for both gene expression levels and PPI connectivity ($p = 0.003$, Table 2). The relative contribution of Hsp90 interaction score and gene expression levels to kinase evolutionary rate was comparable when controlling for the respective other variables; the relative contribution of PPI connectivity was marginal (Table 2).

To quantify the combined explanatory power of Hsp90 interaction score, gene expression levels, and PPI connectivity to kinase dN/dS, we modelled dN/dS PIC on the PIC of all three variables. This model explained only a small component of the dN/dS PIC (R^2= 0.11).

In summary, we found that 1) Hsp90 client status is positively associated with kinase dN/dS, 2) this association appears to be independent of gene expression levels or PPI connectivity, and 3) the strength of association is comparable for Hsp90 client status and gene expression levels. Notably, none of the tested factors alone or in combination explained a large proportion of the observed variation in kinase evolutionary rate.

Hsp90-associated effects on kinase evolutionary rate increase with increasing divergence

Thus far, we have analyzed kinase dN/dS values that were calculated from human and mouse, species which diverged ~75 million years ago (Chinwalla 2002). As mice are known to be a long-branching clade (Wu & Li 1985; Gibbs et al. 2004), we evaluated whether Hsp90's effect on kinase dN/dS values was lineage- or time-dependent. To do so, we obtained pairwise dN/dS values for the orthologs of strong human Hsp90 kinase clients and non-clients from 15 mammalian species. Consistent with the mouse-human comparisons, strong Hsp90 clients evolved faster than non-clients across these species, including several monophyletic lineages such as Primates, Glires, and Laurasitheria (Figure 1b).

Across all comparisons, time of divergence was positively correlated with the ratio of strong client dN/dS to non-client dN/dS (Figure 1c) (R^2 = 0.1783, p = 1.815e-05, Pearson correlation test). This trend was robust to the omission of the fast-evolving rodent lineage (R^2 = 0.2533, p = 5.475e-07, Pearson correlation test; Figure S1a) and was observed even when only

primates were considered (R^2 = 0.318, p-value = 0.01673, Pearson correlation test; Figure S1b). These correlations, however, were not made from independent observations. In addition to the possibility that the evolutionary rates of these kinases are not independent, this analysis may be confounded by dependencies of divergence times between species. Therefore, we accounted for the phylogenetic relatedness among the tested kinases and species simultaneously using a Mantel test, which detects associations between pairwise observations while controlling for their non-independence. We obtained the same result as observed before correction (Mantel test, R^2 = 0.187, p-value = 0.002). Our results suggest that over evolutionary time, Hsp90 client kinases tend to accumulate increasing amounts of non-synonymous variation compared to non-client kinases.

Several previous studies have demonstrated that Hsp90-dependent standing variation is common in yeast, plant, fly, and fish populations (Rutherford & Lindquist 1998; Queitsch et al. 2002; Yeyati et al. 2007; T Sangster et al. 2008; Jarosz & Lindquist 2010). Using genetic or pharmaceutical perturbation of Hsp90, these studies found significantly increased trait heritability upon Hsp90 inhibition, presumably due to the many identified Hsp90-responsive loci. For the vast majority of these Hsp90-dependent loci the identity of the underlying polymorphism (*i.e.* the gene or regulatory region affected) remains unknown and hence evolutionary rate analysis of these loci has yet to be conducted. However, a detailed genetic study found that an Hsp90 client showed increased accumulation of non-synonymous variation compared to its non-client paralog within *A. thaliana* across divergent strains (Lachowiec et al. 2013).

We therefore hypothesized that Hsp90's effect on kinase evolutionary rate should be detectable among humans. To test this hypothesis, we took advantage of the thousands of sequenced human genomes and examined nucleotide diversity of kinase clients and non-clients as a measure of accumulated genetic variation. Specifically, we assessed nucleotide diversity for kinase clients and non-clients in approximately 6500 individuals (Exome Variant Server), accounting for relatedness and using Hsp90 interaction score contrasts as measure of Hsp90 client status.

Indeed, the PIC of kinase Hsp90 interaction scores and nucleotide diversity were positively correlated ($p = 0.04965$, $R^2 = 0.025$, Table S1), suggesting that Hsp90 client kinases harbor greater genetic variation than non-client kinases in humans. We then examined whether Hsp90 clients tolerated potentially more damaging variation across humans, using both Genomic Evolutionary Rate Profiling (GERP) scores (Cooper et al. 2005) and Polymorphism Phenotyping (PolyPhen) scores (Adzhubei et al. 2010) as measures of the predicted phenotypic effect of individual single nucleotide variants (SNVs). GERP ascertains the degree of past purifying selection on a site through examination of rejected nucleotide substitutions across homologous sequences (Cooper et al. 2005), whereas PolyPhen categorizes non-synonymous SNVs by likelihood of damaging protein structure and function (Adzhubei et al. 2010). Although there was no significant correlation between GERP and Hsp90 interaction score contrasts ($p = 0.7593$, Table S1), PolyPhen and Hsp90 interaction score contrasts were positively correlated ($p = 0.00987$, $r = 0.042$, Table S1). To summarize, Hsp90 kinase clients appeared to harbor more genetic variation and more damaging mutations than non-clients across many human individuals.

Hsp90 client status tends to be gained and maintained

Our observation that Hsp90's effect on kinase evolutionary rate increases with divergence time suggests that client status is stable over long time periods. This interpretation contrasts with experimental findings that mutating a single amino acid residue can suffice to dramatically alter a protein's dependence on Hsp90 (Citri et al. 2006). Similarly, paralogs or otherwise related proteins that differ only by a few amino acids can differ in Hsp90 client status (Citri et al. 2006; Lachowiec et al. 2013), suggesting that Hsp90 client status can be highly dynamic. To resolve this apparent contradiction, we analyzed the distribution of Hsp90 client status among gene duplicates in the kinase superfamily. Gene duplicates tended to share Hsp90 client status (Figure 2a, χ^2-test, p = 1.837e-05) indicating that client status has phylogenetic signal.

This analysis only considered genes with one paralog, eliminating one-third of the available kinome data. Therefore, we explicitly tested whether there was phylogenetic signal in client status patterns across the entire kinase phylogeny using Pagel's λ (Pagel 1999). Pagel's λ is a statistic that tests whether phylogeny correctly predicts the patterns of covariance among species on a given trait. Here, we estimated this statistic to test whether kinase phylogeny predicts the patterns of covariance among kinases with regard to Hsp90 client status. Assuming that kinases were either Hsp90 clients or non-clients, we compared a model in which λ was computed using the known kinase phylogeny (Manning et al. 2002) (λ = 0.74) to a model assuming no phylogenetic signal (λ = 0, star phylogeny). The former model including phylogenetic signal was a better fit for the observed pattern of client status across the kinase phylogeny than the model without phylogenetic signal (2δ = 8.94, df = 1, p = 0.00279). We

conclude that client status is not randomly distributed, and that more closely related kinases tend to share client status across the whole kinase superfamily.

Having established that client status tends to be shared among related kinases, we next explored the evolutionary paths by which client status is acquired and changes through time. To do so, we estimated the transition rates between client and non-client states along the kinase phylogeny, using maximum likelihood (ML) implemented in the software BayesTraits (Pagel et al. 2004; Pagel 1999). We found that the rate of the transition from non-client to client along the kinase phylogeny was 3.02, compared to the reverse rate of 0.99. Restricting these rates to equal values significantly worsened the model fit ($2\delta = 8.066$, $df = 1$, $p=0.0045$), hence we conclude that kinases are more likely to become Hsp90 clients than to lose client status (Figure 2b). To examine whether these ML-derived rates were robust, we also used the Markov Chain Monte Carlo (MCMC) analysis implemented in BayesTraits to estimate the rates of gain and loss of client status. The rates of transition for non-client to client (3.11) and client to non-client (0.98) indicated that rate estimates were robust (Figure 2b). The tendency of non-client kinases to become clients and of client kinases to remain clients agrees with previous suggestions that kinase dependence on Hsp90 may be 'addictive' and may contribute to the greatly increased sensitivity of cancer cells to Hsp90 inhibitors (Workman et al. 2007).

If protein kinases indeed become 'addicted' to being Hsp90 clients, one may expect the ancestral kinase to be a non-client. To infer the client status of the ancestral kinase, we compared a tree with a non-client kinase root to a tree with a client kinase root using MCMC. We found strong support for the tree with a non-client root (Bayes factor ~9.1). Both Hsp90 and kinases were present in the eukaryotic common ancestor (Manning & Hunter 2009; Bogumil et al. 2014);

however, our results would suggest that they likely did not interact. Although, Hsp90 client predictions and confirmed clients in the prokaryote *E. coli* include kinases (Press et al. 2013), these prokaryotic protein kinases belong to a different kinase family and differ in structure (Manning & Hunter 2009). Taken together, our results are consistent with a scenario in which early kinases evolved independently of Hsp90 with subsequent multiple independent gains of Hsp90 client status, increasing Hsp90's effect on kinase evolution and adding to the overall chaperone dependence of kinases for their function.

Discussion

Although previous studies showed that Hsp90-dependent standing variation is common in natural populations (Rutherford & Lindquist 1998; Queitsch et al. 2002; T Sangster et al. 2008; Jarosz et al. 2010; Rohner et al. 2013), the chaperone's impact on protein evolution, especially in comparison to established factors such as gene expression, has remained unknown. Here, we present evidence for a direct role of Hsp90 in kinase evolution and compare its impact to gene expression levels and PPI connectivity.

We find that interaction with Hsp90 is associated with higher dN/dS for kinases and that Hsp90 client status appears to contribute to kinase evolutionary rate independently of gene expression and PPI connectivity. Each of these factors contributed to a similar degree to kinase evolutionary rate; combined they explained only about 10% of kinase dN/dS variation. Adding additional variables may decrease the amount of unexplained variation, but many variables are known to co-vary. For example, Bloom and Adami (2003) have argued that PPI connectivity is confounded by protein abundance, with highly abundant proteins engaging in a larger number of

interactions (Bloom & Adami 2003). Protein abundance arises as a combination of gene expression levels, and rates of translation and protein degradation. Others have argued that the dominant role of gene expression in driving evolutionary rate is at least in part due to selection for translational robustness, which encompasses selection for increased translational accuracy by optimizing codon usage, and selection to increase the number of proteins that fold properly despite mistranslation (Drummond et al. 2005). The kinase Hsp90 clients analyzed here are generally less abundant in steady-state protein levels (Taipale et al. 2012), presumably due to their enhanced structural lability. Nevertheless, by using partial contrasts, we find evidence for a contribution of Hsp90 to kinase dN/dS independent of expression level and PPI degree.

As suggested by prior studies of Hsp90-dependent variation within diverse populations of plants, yeast, fly and fish (Rutherford & Lindquist 1998; Queitsch et al. 2002; Yeyati et al. 2007; Todd A Sangster et al. 2008; Jarosz & Lindquist 2010), we found support for Hsp90-dependent genetic variation across humans. Both nucleotide diversity and PolyPhen scores were significantly correlated with kinase Hsp90 interaction scores; yet effect sizes were modest. No significant correlations were observed for GERP scores. The GERP score is a position-specific estimate of evolutionary constraint using maximum likelihood evolutionary rate estimation (Cooper et al. 2005). High evolutionary constraint is often interpreted as functional relevance. As we do not expect Hsp90-dependent variation to reside in highly constrained sites, it is not surprising that GERP scores for Hsp90 kinase clients and non-clients did not differ significantly.

In agreement with the modest signal of increased nucleotide diversity within human, we observed an increased evolutionary rate for Hsp90 client kinases across divergent mammalian lineages, consistent with Hsp90 allowing for greater accumulation of non-synonymous changes

in the genes encoding its clients. In fact, Hsp90's effect on kinase evolutionary rate increases when considering species that are more distantly diverged. The greater accumulation of non-synonymous changes may allow client kinases to explore a wider sequence space and potentially acquire novel functions at a faster rate.

Taipale et al (2012) also determined Hsp90 interaction scores for a large number of transcription factors and E3 ligases. In contrast to our findings for client and non-client kinases, transcription factor and E3 ligase Hsp90 interaction scores were not significantly associated with their evolutionary rates (Text S1, Figure S2). Unlike the kinases analyzed, the transcription factors are not monophyletic (Vaquerizas et al. 2009) and hence may differ more in other factors influencing evolutionary rate, such as protein structure and stability. As for the E3 ligases, others have suggested that Hsp90 works in concert with E3 ligases to promote proteasome-dependent degradation rather than chaperoning these enzymes (Murata et al. 2001; McClellan et al. 2005; Morishima et al. 2008; Ehrlich et al. 2009; Taipale et al. 2010). Hsp90 works in concert with many other proteins that enable and modify its function and generate client specificity, such as diverse Hsp70s, co-chaperones, and immunophilins (Taipale et al. 2010). Although these "collaborating" proteins physically interact with Hsp90, they do so in a sequence-or domain-specific manner, and Hsp90 does not facilitate their folding (Murata et al. 2001; McClellan et al. 2005; Morishima et al. 2008; Ehrlich et al. 2009; Taipale et al. 2010); hence they are unlikely to experience relaxed selection due to this interaction. Following this line of reasoning, we previously excluded known Hsp90 co-chaperones and Hsp70s from evolutionary rate analyses (Lachowiec et al. 2013).

Previously, we found that Hsp90 clients in plants and yeast showed significantly greater evolutionary rates than their non-client paralogs (Lachowiec et al. 2013). The kinome data set only contained five pairs of strong kinase clients and non-clients. Although, in four of these pairs the strong kinase client showed greater dN/dS than its non-client paralog, this difference was not significant, presumably due to small sample size (n=5, 95% confidence interval 0.6-12.6, p=0.125, one-sample Wilcoxon test, testing the deviation from the expected ratio of client/non-client dN/dS of 1). We did, however, detect a significant trend for duplicate genes such that paralog pairs that contained at least one strong Hsp90 client showed significantly greater divergence than pairs that did not (Figure S3). We speculate that the interaction with Hsp90 allows both enhanced neo-functionalization through tolerated non-synonymous variation and sub-functionalization through the client's dependence on the environmentally-sensitive chaperone (Lachowiec et al. 2013).

Beyond paralog pairs, we further explored the relationship of Hsp90 client status and kinase evolutionary rate among kinase families. The effect of Hsp90 client status on dN/dS was consistent across all kinase groups (Figure S4a) with greater dN/dS observed for Hsp90 client kinases compared to non-client kinases. The TK and TKL families were enriched for strong Hsp90 clients (Figure S4b), and possibly in part due to this enrichment, both families showed the highest evolutionary rates among the tested kinase families (Figure S4c). TK and TKL kinase families are evolutionarily young and have been implicated in the rise of multicellularity (Lim & Pawson 2010). In contrast, of all families, the CAMK family was most depleted for strong Hsp90 clients. Unlike in all other kinase families, the few strong CAMK clients did not evolve faster than the weak clients (Figure S5). As almost a third of the CAMK family genes are pseudogenes

(not included in this analysis), contrasting with only about one-fifth of kinase pseudogenes overall (Manning et al. 2002), we speculate that the "missing" strong CAMK clients have become pseudogenes.

These family-specific observations and our finding that kinases tend to acquire and maintain Hsp90 client status prompt our speculation that Hsp90 may play a complex role in the birth and death of kinases. The most common outcome after gene duplication is pseudogenization of one copy (Nei & Roychoudhury 1973). Acquiring Hsp90 client status likely leads to instant sub-functionalization due to the temperature sensitivity of clients and may facilitate gene copy maintenance (Lachowiec et al. 2013). The observed greater accumulation of non-synonymous variation in genes encoding Hsp90 client may also facilitate neo-functionalization. At the same time, the greater accumulation of more harmful variation may predispose genes encoding clients to the pseudogene fate. In other words, acquiring Hsp90 client status may be akin to a delayed death sentence on a long evolutionary time scale.

In fact, the suggested "Hsp90 addiction" of kinases (Workman et al. 2007), especially in cancer cells with their mutated oncogenic kinases, is reminiscent of a recent argument by Fernandez and Lynch (Fernández & Lynch 2011). These authors attributed the increasing complexity of protein interaction networks from bacteria to human to compensation for the decreased stability of proteins over evolutionary time. They posited that through drift proteins would be exposed to destabilizing mutations and hence become susceptible to aggregation and malfunction. Interaction in homo-and hetero protein complexes will then compensate for the stability deficits of proteins in higher organisms (Fernández & Lynch 2011). Of course, chaperones also prevent aggregation and stabilize proteins (Taipale et al. 2010). The small

evolutionary snapshot of the kinome with its tendency to acquire and maintain Hsp90 client status fits well within this framework of thought.

Methods

Data sources and estimating contributions to dN/dS

Hsp90 interaction scores, client category, and number of connections in the protein-protein interaction (PPI) network were obtained from (Taipale et al. 2012). For gene expression levels in human, RNA-seq counts normalized for read depth (RPKM, Reads per kilo-base per million) across eleven normal human tissues from (Castle et al. 2010) were downloaded (http://medicalgenomics.org/rna_seq_atlas/ [September, 2012]). The average expression was taken per gene (across all splice forms) and the maximum and mean expression across all tissues was calculated. Only the 210 kinases that overlapped between the Hsp90 kinome analysis (Taipale et al. 2012) and the eukaryotic kinase tree (Manning et al. 2002) and that had values for expression (Castle et al. 2010), dN/dS, and PPI connectivity were used in the downstream analyses. Since the phylogenetic relatedness of the kinases may influence these contributions, we also considered each variable in a phylogenetic context using phylogenetically independent contrasts (PIC). Phylogenetic contrasts for each variable were estimated in R using *pic* in the package ape (Paradis et al. 2004) with the kinome tree from (Manning et al. 2002) (http://kinase.com/human/kinome/groups/ePK.ph). We conducted linear regression analyses with the intercept set to 0 (Garland et al. 1992) since the order of subtraction to calculate the PIC was arbitrary and calculated the association between each pair of variables. Partial correlations

among the variables were calculated with R using *pcor* in the package ppcor (http://cran.r-project.org/web/packages/ppcor/index.html).

dN/dS comparisons across species

For each kinase, orthologs for the human kinases and their respective dN/dS values were identified using Ensembl release 70, (http://www.ensembl.org [March, 2013]) in the following species: *Pan troglodytes, Pongo abelii, Nomascus leucogenys, Macaca mulatta, Callithrix jacchus, Mus musculus, Rattus norvegicus, Ictidomys tridecmlineatus, Oryctolagus caniculis, Canis familiaris, Bos taurus, Sus scrofa,* and *Ailuropoda melanoleuca*. If multiple orthologs were identified, the kinase was removed from the analysis. Time to common ancestor was obtained from compiled studies curated at TimeTree.org (Hedges et al. 2006).

Hsp90 client divergence timing among humans

We acquired human exome sequences from the Exome Variant Server (Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (http://evs.gs.washington.edu/EVS/ [November, 2012]) using an in-house Perl script.

In addition to the filters from the ESP Server, we incorporated the filters used in (Tennessen et al. 2012): all SNVs had to have a quality score above 20, allele balancing above 65%, read depth between 10 and 1000x to account for possible copy variation, and we excluded genotype qualities of zero.

To calculate nucleotide diversity, we used

$$\pi_{SNV} = 2f(1-f)\frac{n}{n-1}$$

where f is the frequency of the major allele and n is the number of haploid genomes (Hernandez et al. 2011). For each gene π_{SNV} was summed for each nucleotide position and normalized for gene length. For gene length we used the length of the longest CDS reported by Ensembl release 69.

We extracted GERP scores for each SNV from ESP and calculated the average GERP score for each kinase gene and normalized the GERP counts by the length of the gene (CDS sequence). PolyPhen scores were also obtained from ESP.

Examining client status dynamics and estimating transition rates between client and non-client states

We estimated the phylogenetic signal of client status across the kinome, Pagel's λ, using the function *fitDiscrete* in the package geiger (http://cran.r-project.org/web/packages/geiger/index.html). We modeled a kinase tree with all branches leading to tips of equal length (λ = 0) eliminating the phylogenetic signal. We then compared this model to a model with the true kinase phylogeny (i.e. an optimized λ) using likelihood-ratio tests, where the likelihood ratio approximates a χ^2 distribution with one degree of freedom (Harmon et al. 2008; Motani & Schmitz 2011).

We used BayesTraits v2 (Pagel 1994) (http://www.evolution.rdg.ac.uk/BayesTraits.html) to estimate transition rates along the kinase tree using a MultiState model of evolution. We used the human kinase tree and the client status of each kinase, coding both weak and strong clients

(according to Taipale et al, 2012) as ‘A’ and non-clients as ‘B’. We then used maximum likelihood with the parameter rate set to 2, representing the two transition rates 1) client to non-client and 2) non-client to client. We tested various hypotheses about transition rate parameters by restricting transition rates: 1) rate of gain of client status to 0, 2) rate of loss of client status to 0, and 3) equal gain and loss rates. We compared the log(likelihood) of the restricted models to one another and to the log(likelihood) of the unrestricted models. For these comparisons we used likelihood-ratio tests, where the likelihood ratio follows a χ^2 distribution with degrees of freedom equal to the number of restricted parameters. We repeated the analysis using Markov Chain Monte Carlo (MCMC) implemented in BayesTraits v2. We had appropriate levels of acceptance with the option rateDev set to 2.We let the chain run for 100,000 iterations with a uniform prior distribution between 0 and 100. We examined the transition rates after a 20,000 iteration burn-in period.

To determine the client status of the root kinase, we used MCMC implemented in BayesTraits v2. We used the kinase tree and client status coding described above. We fossilized the root as either A or B for the whole tree. We found appropriate levels of acceptance with the option rateDev set to 1. We ran the chain for 10 million iterations, and compared the two different client states of the root by calculating the Bayes Factor based on the harmonic means (twice the difference between the two harmonic means of the model likelihood) (Pagel et al. 2004).

Examining non-kinase Hsp90 interactors

We removed all genes that were found in more than one functional category: E3 ligases, transcription factor, or kinase. The client status as defined in Taipale et al (2012) was used for each gene, with E3 ligases and transcription factors (TFs) categorized as "not significant interactor", removed from further analyses. We used pairwise dN/dS values with mouse as an outgroup to examine the dN/dS among the Hsp90 interactors and non-interactors. Neither TFs nor E3 ligases are monophyletic. We subdivided the TFs into phylogenetically related families (Vaquerizas et al. 2009) and compared the dN/dS between Hsp90 interactors and non-interactors within each family and in aggregate. Subdividing E3 ligases into phylogenetic groups based on sequence similarities has not been previously conducted, so we subdivided E3 ligases based on domain presence (Kelch, WD40 from (Taipale et al. 2012) or RING, U box, HECT, F box, SOCS box, BTB, DDB1-like, ZnF A20 from (Li et al. 2008)) and compared the dN/dS between Hsp90 interactors and non-interactors within each group. Because no differences in dN/dS were found between Hsp90 interactors and non-interactors within groups defined by individual domains, we also clustered E3 ligases based on presence or absence of many domains simultaneously. E3 domains were identified using PFAM version 26. Clustering was completed using hierarchical clustering in R with the function *heatmap*.

Acknowledgments

We thank Maximilian Press and Mikko Taipale for helpful discussions. We thank the NHLBI GO Exome Sequencing Project and its ongoing studies which produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the WHI Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926) and the Heart GO Sequencing Project (HL-103010). This

work was supported by the National Science Foundation (DGE-1256082 to JL), the National Institutes of Health (DP2OD008371 to CQ, DP2AT00780201 to EB), and the Alfred P Sloan Foundation (to EB).

Main Figures

Table 1 Hsp90 interaction score is positively associated with dN/dS.

Model	regression coefficient		R^2
PIC[1] dN/dS ~ PIC Hsp90 interaction score	0.007462	***	0.06447
PIC Hsp90 interaction score ~PIC expression maximum[2]	-0.003507		-0.004182
PIC dN/dS ~ PIC expression maximum	-0.0006343	*	0.02056
PIC Hsp90 interaction score ~ PIC PPI	-0.35586	****	0.08779
PIC dN/dS ~ PIC PPI	-0.008626	***	0.06256

[1]phylogenetic independent contrasts
[2]maximum expression across 11 tissues
p-values:*0.05, **0.01, ***0.001, ****0.0001

Table 2. Hsp90 is correlated with dN/dS when controlling for PPI and expression[1].

	PIC[5] HIS[2]	PIC dN/dS	PIC PPI[3]
PIC HIS[2]			
PIC dN/dS	0.197**		
PIC PPI	-0.105	-0.0958	
PIC max expr[3]	-0.139*	-0.212**	0.027063

[1]Partial correlation are shown between two variables, controlling for the other two
[2]HIS- Hsp90 interaction score
[3]PPI-number of protein-protein interactions
[4]max expr – maximum expression across 11 tissues
[5]Phylogenetic independent contrasts
p-values:*0.05, **0.01

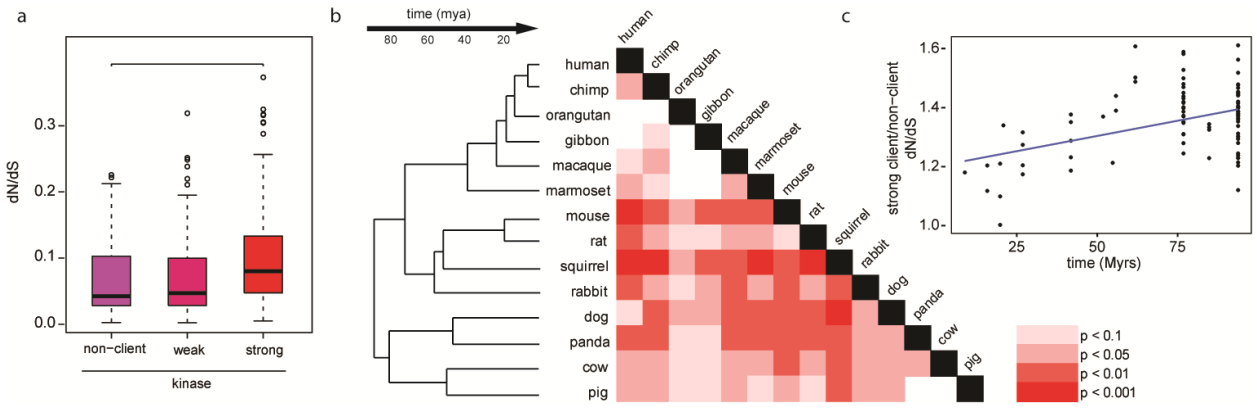

Figure 1. Hsp90 client and non-client kinases differ significantly in evolutionary rate. a) The difference for human-mouse pairwise dN/dS for client and non-client kinases is driven by strong clients ($p = 0.0004211$, Wilcoxon rank-sum test). b) Significant differences between strong and non-client kinase dN/dS were observed across mammals (Wilcoxon rank-sum test). Shades of red indicate significance levels. c) The ratio of the average strong client dN/dS and average non-client dN/dS increases as the time to common ancestor increases between species (linear regression, $R^2 = 0.1783$, $p = 1.815e-05$). This trend is robust to the omission of rodents (linear regression, $R^2 = 0.2533$, $p = 5.475e-07$, Figure S1a) and is observed within primates alone (linear regression, $R^2 = 0.318$, p-value = 0.01673, Figure S1b).

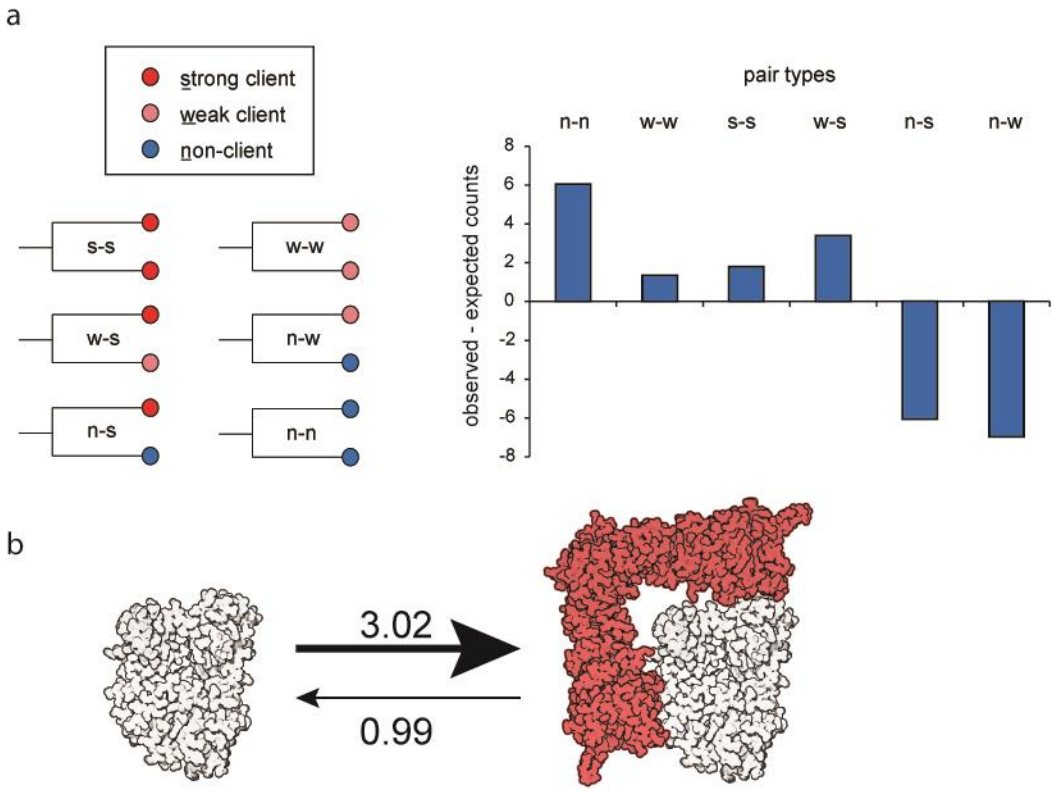

Figure 2. Hsp90 client status dynamics across the kinome. a) Gene duplicates tend to share client status (χ^2-test, p = 1.837e-05). The number of expected pairs was calculated based a random distribution of clients across the tree. b) Kinases (white) are three times more likely to gain Hsp90 client status (red) than lose it based on ML and MCMC estimates of state transition rates using Bayes Traits.

Supplemental Figures

Table S1. Hsp90 client kinases harbor more and more harmful genetic variation across humans.

	p-value	Rho (Pearson's)
PIC[1] nucleotide diversity ~ PIC HIS[2]	0.04965	0.157
PIC PolyPhen probable ~ PIC HIS	0.0099	0.206
PIC GERP ≥ 5 ~ PIC HIS	0.7593	0.02473

[1]Phylogenetic independent contrasts
[2]HIS- Hsp90 interaction score

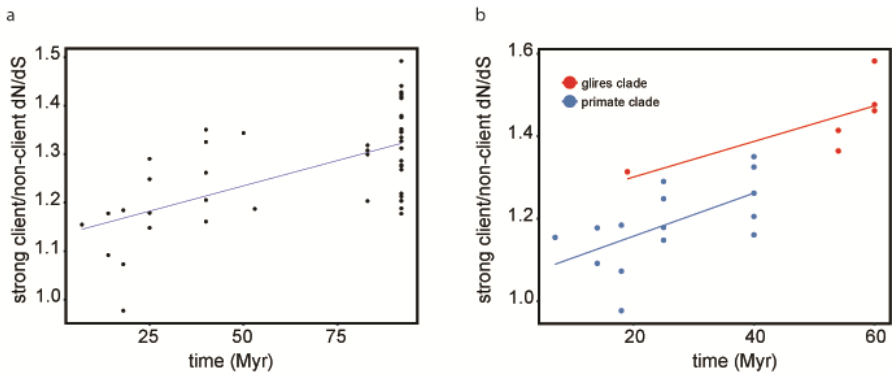

Figure S1. The ratio of average strong client kinase dN/dS to non-client kinase dN/dS increases across different clades of mammals. a) This trend is robust to removing the fast-evolving Glires (rabbits and rodents, linear regression, p=5.475e-07 R^2=0.2533). b) The same trend is observed considering smaller, more recent clades such as primates only (primates: linear regression, p=0.01673, R^2=0.318).

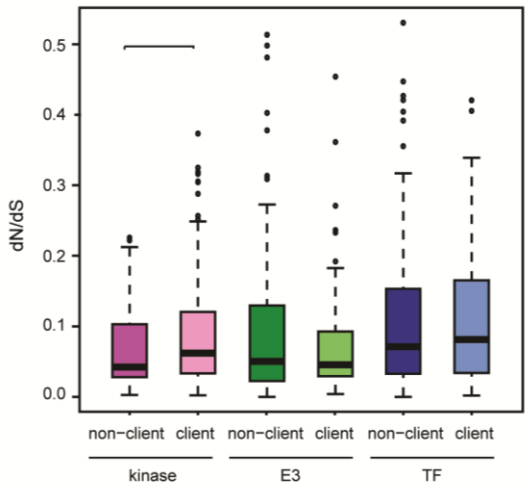

Figure S2. E3 and TF clients do not evolve faster than their respective non-clients. In contrast to kinases (p = 0.01221, Wilcoxon rank-sum test), transcription factor and E3 ligase clients do not show significantly greater dN/dS than transcription factors and E3 ligase non-clients.

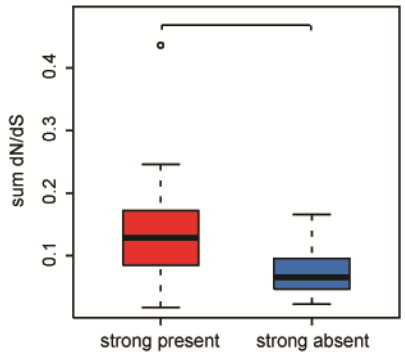

Figure S3. Hsp90 client status is associated with divergence of gene duplicates. Divergence was measured as the sum dN/dS for each gene duplicate pair. Pairs of gene duplicates that encode at least one strong Hsp90 client (n=19) diverge faster than pairs without an Hsp90 client (n=27) ($p = 0.001183$, Wilcoxon rank-sum test).

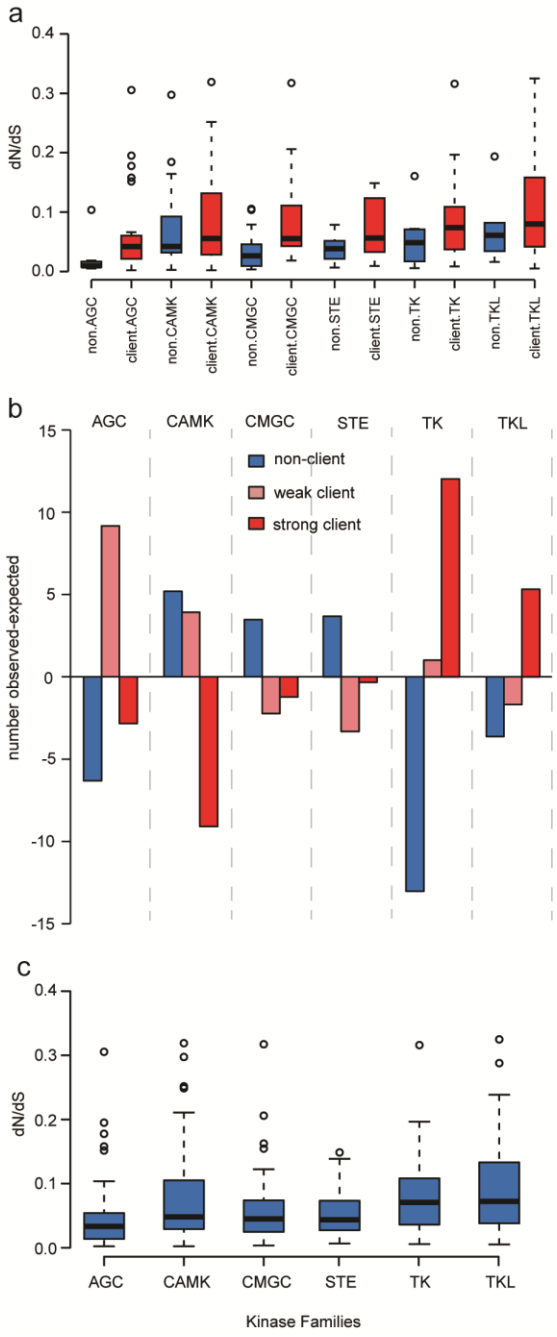
a
dN/dS
0.0
0.1
0.2
0.3
0.4
non.AGC
client.AGC
non.CAMK
client.CAMK
non.CMGC
client.CMGC
non.STE
client.STE
non.TK
client.TK
non.TKL
client.TKL
b
AGC
CAMK
CMGC
STE
TK
TKL
non-client
weak client
strong client
number observed-expected
15
10
5
0
-5
-10
-15
c
dN/dS
0.0
0.1
0.2
0.3
0.4
AGC
CAMK
CMGC
STE
TK
TKL
Kinase Families

Figure S4. Hsp90's effect on dN/dS is observed in all kinase groups. a) In each kinase group, the Hsp90 clients tended to show greater dN/dS than non-clients. b) The TK and TKL groups were enriched for strong Hsp90 clients; the CAMK group was most depleted for strong Hsp90 clients. The expected number of non-clients, weak clients, and strong clients was calculated by assuming an equal distribution of all client states across all kinase groups. c) The TK and TKL families show the largest median values of dN/dS, albeit their dN/dS did not significantly differ from the other kinase groups.

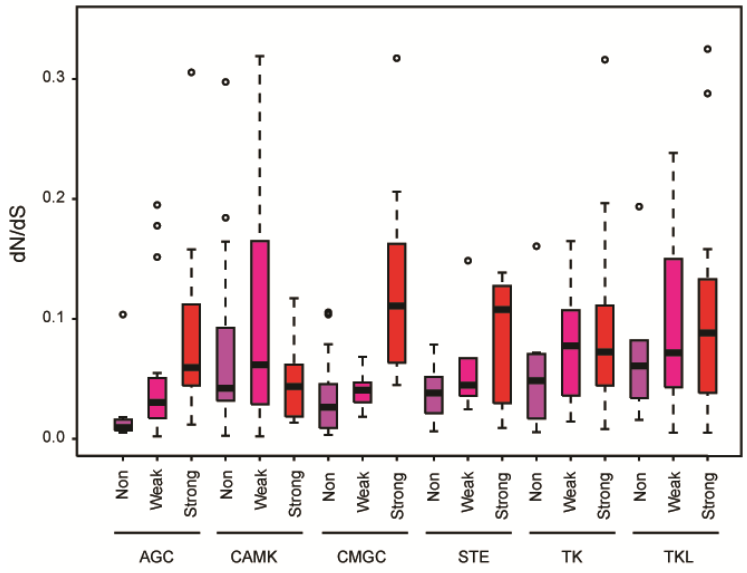

Figure S5. Family-specific differences for the association of Hsp90 client status and dN/dS. In general, strong kinase clients (red) tended to have greater dN/dS than weak kinase clients (pink), and weak kinase clients (purple) tended to have greater dN/dS than kinase non-clients. The TK and CAMK groups did not follow this pattern, with weak kinase clients tending to have greater dN/dS, presumably due the broad distribution of dN/dS for weak kinases in these groups.

Supplementary Text

Interaction with Hsp90 is associated with faster evolutionary rates in clients, but not co-chaperones or otherwise collaborating proteins.

We failed to detect an effect of Hsp90 client status on the evolutionary rate of transcription factors and E3 ligases (Figure S2). In contrast to the tested kinases, which evolved from a common ancestor, transcription factors and E3 ligases are not monophyletic (Li et al. 2008; Vaquerizas et al. 2009) and hence likely differ greater in other features influencing evolutionary rate. Hence, we subdivided the tested transcription factors (Vaquerizas et al. 2009) and E3 ligases (Li et al. 2008) into phylogenetically related groups for further tests.

For transcription factors, this subdivision approach was hampered by the extremely low number of transcription factors that interact with Hsp90 (58 out of 843 tested) (Taipale et al. 2012). Comparing evolutionary rates of clients and non-clients in phylogenetically related groups was not feasible (average group size 18.4, average number of clients 1.8, and non-clients 16.7).

For E3 ligases, we used previously described subdivisions and tested for association between Hsp90 client status and dN/dS. Specifically, we divided the E3 ligases into RING finger domain-containing (RNF) and non-RNF proteins (Li et al. 2008). We found no significant effect of Hsp90 client status on RNF and non-RNF E3 ligases. Similarly, testing only E3 ligases with a Kelch fold (Taipale et al. 2012) yielded no evidence that Hsp90 interactors evolved faster than non-interactors. E3 ligases also contain small domains with roles in recognition of ubiquitinated substrates. We attempted to cluster the tested E3s by these domains as identified in Pfam, using

several different thresholds for domain identity. Again, we found no significant differences in dN/dS associated with Hsp90 interaction in these groups.

This persistent failure in light of the clear and consistent signal in kinases prompted us to further explore the literature on Hsp90 and E3 ligase interaction. In contrast to TFs and kinases, which require Hsp90's assistance to reach their mature fold, E3 ligases may interact physically with Hsp90 for another reason. Indeed, some E3 ligases have been shown to collaborate with Hsp90 in the degradation of other proteins, rather than being chaperoned by Hsp90 (Murata et al. 2001; Giannini & Bijlmakers 2004; Morishima et al. 2008). If E3 ligases generally are Hsp90-collaborating proteins, we would not expect an Hsp90-associated effect on E3 ligase dN/dS, akin to prior observations with other Hsp90-associated proteins such as Hsp70 and various Hsp90 co-chaperones, all of which are rather highly conserved across eukaryotes.

CHAPTER IV

FINE-MAPPING AN HSP90-BUFFERED LOCUS AFFECTING HYPOCOTYL LENGTH

Introduction

Not all genetic variation translates into phenotypic variation. Evolution allows for genetic variants that do not affect fitness, commonly called standing genetic variation, to accumulate and drift (Paaby & Rockman 2014). However, under certain conditions, this neutral genetic variation may become beneficial or deleterious. Cryptic genetic variation refers to previously silent genetic variation that affects phenotypes when conditions change (Paaby & Rockman 2014). Evidence for cryptic genetic variation influencing phenotypes abounds (Gibson & Dworkin 2004). For example, populations of the threespine stickleback thriving in saltwater have colonized freshwater environments a number of times in the past. Body size variation is low in saltwater dwelling stickleback, but exposure to freshwater increases variance in body size, through a revelation of cryptic genetic variation (McGuigan et al. 2011).

Protein chaperones, which facilitate the folding of protein substrates and are environmentally regulated, are prime candidate genes that can promote the both the tolerance of genetic variation and its release (Tokuriki & Tawfik 2009; Williams & Fares 2010; Bogumil & Dagan 2010; Bogumil et al. 2012; Sangster et al. 2004). The conserved eukaryotic protein chaperone Hsp90 is the best-studied gene whose environmental regulation both reveals and conceals genetic variation (Rutherford & Lindquist 1998; Queitsch et al. 2002; Yeyati et al. 2007; Cowen & Lindquist 2005). The expression of Hsp90 increases in any environment change

that affects protein stability, from changes in temperature, pH, or salinity to the presence of toxins and heavy metals (Stensløkken et al. 2010; Park et al. 2010; Cid et al. 2010; Pratt et al. 2010; Sekimoto et al. 2010; Gerspacher et al. 2009). The ubiquity, promiscuity, and responsiveness of Hsp90 argue that it has an impact on cryptic genetic variation.

Studies across species have detected cryptic genetic variation relying on Hsp90 for masking. Reduction of Hsp90 function results in the expression of line-dependent phenotypes in *D.melanogaster* (Rutherford & Lindquist 1998), *A. thaliana* (Queitsch et al. 2002), *S. cerevisiae* (Jarosz & Lindquist 2010), and both cavefish (Rohner et al. 2013) and zebrafish (Yeyati et al. 2007). In plants and yeast, genetic variation dependent on Hsp90 function was mapped (T Sangster et al. 2008; Jarosz & Lindquist 2010) using recombinant inbred lines (RILs). Jarosz & Lindquist (2010) phenotyped RILs on a control medium and a medium containing an Hsp90 inhibitor. Hsp90 buffered-loci were defined as loci that were mapped only in the Hsp90-inhibited conditions. One-hundred two of the 104 traits examined mapped to an Hsp90-buffered locus in both species (T Sangster et al. 2008; Jarosz & Lindquist 2010). Four of those Hsp90-buffered loci were fine-mapped in *S. cerevisiae* to the causal variant. Three of these four loci map to open reading frames, of which two encode known Hsp90 substrates. The fourth locus mapped to a 3' untranslated region (UTR) demonstrating that Hsp90 can indirectly buffer non-coding genetic variation, presumably through interaction with a protein that recognizes the 3'UTR (Sangster et al. 2004; Jarosz & Lindquist 2010). Indeed the expression level of the gene encoding the 3'UTR Hsp90-dependent variation was increased more than 100-fold in the presence of the Hsp90 inhibitor (Jarosz & Lindquist 2010). From these studies we know that cryptic genetic variation buffered by Hsp90 is common.

In *A. thaliana*, none of the loci buffered by Hsp90 have been fine-mapped to candidate genes or the underlying genetic variant. Sangster and colleagues performed QTL mapping for hypocotyl length and identified a locus on chromosome 2 dependent on Hsp90 in three different mapping populations (Todd A Sangster et al. 2008; T Sangster et al. 2008). Near isogenic lines (NILs) narrowed this region to 10Mb (T Sangster et al. 2008), and Sangster (2007) further completed association mapping across 60 accessions and identified a 30kb region with significant association. This region encodes five candidate genes and phenotyping of T-DNA insertion lines of these five candidate genes implicated two as likely encoding the causal variation (Sangster 2007). We continued this study to fine map the Hsp90-buffered genetic variation on chromosome 2.

Results

Association mapping of Hsp90-buffered locus narrows region to a 30kb containing five genes

Sangster (2007) performed association mapping using a panel of 60 *A. thaliana* accessions to narrow the genomic region containing an Hsp90-buffered locus previously identified using genetic linkage (T Sangster et al. 2008). The accessions exhibited a range of hypocotyl lengths and responses to Hsp90 inhibition (Figure 1) (T Sangster et al. 2008). The two parental lines for the QTL and NIL analysis, CVI and L*er* were included in this panel. CVI was among the least responsive accessions, whereas L*er* was more responsive to Hsp90 inhibition (Figure 1) (T Sangster et al. 2008). Untreated hypocotyl length was correlated to Hsp90 inhibited hypocotyl length across the accessions, indicating that not all natural variation is Hsp90 responsive ($R^2 = 0.43$, $p = 7.2*10^{-7}$, Spearman's rank test).

Twenty to thirty accessions with the most extreme response to Hsp90 inhibition were Sanger sequenced every 3kb between 9.55 and 9.72Mb on chromosome 2 (Sangster 2007). With this approach, a significantly associated 30kb region was identified, which encompassed five candidate genes (Figure 2) (Sangster 2007). In the time since this work was completed, many of the Sanger sequenced accessions have been densely genotyped at an average of every 500bp, possibly allowing for more precision in association. Fifty-six of the accessions tested by Sangster were included in the genotyped panel (Table S1) (Atwell et al. 2010). We performed an association analysis[1] across these 56 accessions from 9.5Mb to 9.75Mb on chromosome 2, only considering SNPs with the minor allele present in at least 6 of the 56 accessions. No significant associations were detected after multiple testing correction (Figure S1a). Sangster's approach using the most extreme phenotypes identified an associated locus, so we completed a similar analysis using the top and bottom quartiles of Hsp90-responsive accessions. Associations were tested with 28 accessions across 102 SNPs (Wilcoxon rank sum test), but again no associations were found after the multiple testing correction (Figure S1b). We conclude that using Sanger sequencing did identify the associated sub-region, so we continued studies on the five candidate genes.

Examining T-DNA insertion lines in five candidate genes for hypocotyl phenotypes

The 30kb region identified by both Sangster's association and linkage studies (T Sangster et al. 2008; Sangster 2007) encodes five protein-coding genes: At2g22460, encoding a protein of unknown function; At2g22470, encoding the arabinogalactan protein AGP2; At2g22475, encoding a GL-2 expression modulator (GEM); At2g22480, encoding a phosphofructokinase

[1] Thanks to Örjan Carlborg and his lab for providing the filtered database of genotyped SNPs.

PFK5; and At2g22490, encoding a sucrose-responsive D-type cyclin (Swarbreck et al. 2008). Bacterial transfer-DNA (T-DNA) insertion lines affecting each of these genes were obtained. Expression of each of these genes was reduced among these T-DNA mutants, except for At2g22480 (Figure S2). At2g22480 was overexpressed, suggesting feedback regulation of the gene in response to its knock-out. Using an additional primer set, we found that the 3' end of At2g22480 was not expressed indicating that the T-DNA insertion did disrupt the expression of that part of the gene.

For Hsp90 inhibition experiments, Sangster used a potent and specific inhibitor of Hsp90, geldanamycin (GdA) (McLellan et al. 2007). Sangster previously observed that the response to Hsp90 inhibition was significantly greater than wild-type in two independent insertions of At2g22490 demonstrating that At2g22490 encodes an Hsp90-buffered polymorphism (Sangster 2007). An insertion in At2g22480 exhibited a significantly reduced response to Hsp90 inhibition compared to wild-type. Based on Sangster results, both of these genes encompass Hsp90-dependent polymorphisms, but the mode of buffering by Hsp90 varies between the two genes.

Using bulked seeds in assays can prevent false positive results due to stochastic silencing of the T-DNA insertion that spreads beyond the gene of interest (Vaucheret et al. 1998; Brodersen & Voinnet 2006). Indeed Sangster found that neither of the At2g22490 T-DNA insertion lines he tested were kanamycin resistant, suggesting that the T-DNA was silenced and may have spread beyond At2g22490. We sought to replicate Sangster's findings but used bulked seeds rather than seeds from single seed descent. Additionally, Sangster did not examine a T-DNA insertion line for the gene At2g22475, which lies within the associated and linked regions, so we added that line to further analyses. After re-genotyping the T-DNA insertion lines, we

observed the reduced hypocotyl length of the At2g22480 insertion line on control media and its lack of response to Hsp90 inhibition (Tukey's HSD post hoc test, Figure 3). However, we did not observe a significant difference in length in an At2g22490 line on GdA (Tukey's HSD post hoc test, Figure 3). Also, no difference between Col-0 and At2g22475 was observed. These results contradict Sangster's findings that At2g22490 encompasses an Hsp90-buffered polymorphism.

The lack of phenotype for the At2g22490 line on GdA may be due to the dose used. Therefore, we grew the At2g22475, At2g22480, and At2g22490 T-DNA insertion lines at increasing concentrations of GdA. No differences in response to Hsp90 inhibition were observed in the insertion lines At2g22475 or At2g22490 (Figure 4a). The lack of response to GdA was observed for At2g22480 across the three GdA concentrations (Figure 4a). With the dose curve, we again did not detect a difference in the At2g22490 line's response to Hsp90 inhibition, providing further support that it does not encode Hsp90-buffered genetic variation.

At2g22480, *PFK5*, is left as the most likely candidate after eliminating At2g22490 as containing Hsp90-buffered genetic variation. However, it is possible that the phenotype observed is due to a truncated PFK5 because the At2g22480 T-DNA insertion line tested, At2g22480 SAIL, interrupted At2g22480 in the second to last exon, and the portion of the gene upstream of the insertion was expressed (Figure S2). We acquired several additional SALK T-DNA insertion lines for At2g22480 to test if the phenotype was consistent across different mutants. Lines At2g22480 SALK1 and At2g22480 SALK3 interrupted the promoter, whereas At2g22490 SALK2 interrupted the 8^{th} and 9^{th} exons (of 13 exons). None of the three SALK alleles exhibited the short hypocotyl phenotype on control media as the At2g22480 SAIL insertion line did (Figure 4b, $p < 0.05$, Tukey's HSD post-hoc test), indicating that the short hypocotyl phenotype

is specific to At2g22480 SAIL. We suggest that At2g22480 SAIL insertion truncates *PFK5* in a specific manner that affects both hypocotyl length and response to Hsp90.

PFK5 is a phosphofructokinase (PFK), and a member of a small family of seven PFK, *PFK1-7* (Figure 5) in *A. thaliana*. PFKs perform the commitment enzymatic step of glycolysis, the conversion of fructose-6-phosphate to fructose-1,6-bisphosphate to create ATP (Pratt & Cornely 2012). *PFK5* forms its own clade in the *PFK* gene family phylogeny, suggesting that it has diversified in function from the other family members (Figure 5). *PFK5* localizes to the exterior of chloroplasts and is expressed in root, young leaf, mature leaf, stem, flower (Mustroph et al. 2007), and also seedlings (Hruz et al. 2008).

Allele-swap analyses

To confirm whether At2g22480 encodes Hsp90-buffered loci, we performed allele swaps. For At2g22460, At2g22470, At2g22475, At2g22480, and At2g22490, we transformed the respective T-DNA line with the L*er*, or CVI allele. As a control, each T-DNA line was also transformed with the Col-0 of the interrupted gene. These T-DNA lines are in the Col-0 background providing the opportunity to see if the CVI and/or L*er* alleles are causal to the hypocotyl response to Hsp90 inhibition. The constructs include 1000bp upstream and 1000bp downstream of the gene in question so that promoter and UTR variation may be ascertained as well.

To ensure that the lines created are correct, we will amplify each allele from the transgenic lines and sequence. We will then analyze the allele swaps to determine whether the regions encoded by the constructs contain genetic variation buffered by Hsp90. Hypocotyl

lengths of two to three independent lines of each insertion will be examined both on control and GdA media. Each independent line will have the insertion at a different position, allowing for position effect to be controlled.

Discussion and future directions

Sangster and colleagues (2008) had located a 2Mb region containing Hsp90-buffered genetic variation through QTL mapping. Using near isogenic lines narrowed the associated region to 200kb, and association mapping further narrowed the region to 30kb, which encodes only five candidate genes. I attempted to narrow the candidate gene region further using the publically available dense genotyping data for many accessions (Atwell et al. 2010). An average of 500bp is positioned between each genotyped variant. However, no loci were found to be associated with Hsp90 response. I speculate that the loci identified by Sangster (2007) are in a short block of linkage disequilibrium, explaining we were unable to narrow the region of candidate genes further.

We found further support for one of the candidate genes containing Hsp90-buffered genetic variation initially identified by Sangster (2007), At2g22480. The At2g22480 SAIL T-DNA insertion line was insensitive to Hsp90 inhibition, even at increasing doses, suggesting that the T-DNA insertion itself is buffered by Hsp90. At2g22480 encodes a PFK, *PFK5*, which regulates the commitment step in the process of glycolysis (Pratt & Cornely 2012). Interestingly, Clark reports in her thesis that Hsp90 binds PFK1 and PFK2 in human fibroblasts (Clark 2009). This report strongly suggests that PFK5 is an Hsp90 client, providing a mechanism for why *PFK5* contains Hsp90-buffered genetic variation.

Different At2g22480 T-DNA insertion lines had different phenotypes in control conditions, adding a challenge in demonstrating whether At2g22480 does indeed contain variation buffered by Hsp90. The T-DNA insertion lines differ in the position of their insertion, and there is evidence that the At2g22480 was still expressed in At2g22480 SAIL. To explain the observed phenotypes, I have developed a model based on the possible expression levels of the At2g22480 in the T-DNA lines.

In the model, I assume that At2g22480 expressed in the SAIL line, but At2g22480 is not expressed in the line SALK2 (Figure 6). Expression of At2g22480 was up-regulated 3-fold in the SAIL line (Figure S2), so the reduced hypocotyl length may be due to over-expression. If over-expression is associated with reduced length, loss of function may be associated with increased growth. Indeed, the SALK1 and -2 exhibited significantly larger hypocotyls than wild-type. This model suggests that At2g22480 is a negative regulator of hypocotyl growth. I rejected models that assumed an equal or over-expression of SALK2 because they were not consistent with the increased hypocotyl length observed in SALK2. Careful examination of expression of full-length and truncated versions in all of the T-DNA insertion lines would demonstrate whether the proposed model should be rejected.

Another challenge in interpreting the T-DNA results was the overexpression present in the SAIL line. Is the shortened hypocotyl phenotype due to the truncation or overexpression? The longer hypocotyls observed in the SALK lines support that the phenotype was due to expression levels. Creation of an overexpression line for full-length PFK5 and a low expression truncated PFK5 line would demonstrate whether expression level or protein length was causal.

The last difficulty is explaining why a truncated or overexpressed PFK5 is insensitive to Hsp90 inhibition. If PFK5 is a client, one possibility is that the truncated PFK5 is no longer dependent on Hsp90, explaining the insensitivity of the mutant. Alternatively, the truncated PFK5 could continue to be a client, but the overexpression of PFK5 overcomes the requirement for Hsp90. This explanation is possible because client status is a continuum (Taipale et al. 2012), possibly due to the differential requirement of individual proteins for Hsp90. The determination of client status of the full-length and truncation PFK5 would prove which of the possibilities is correct.

Though there are many difficulties in understanding the T-DNA insertion line data, the experiments that will show whether PFK5 harbors genetic variation buffered by Hsp90 are the allele swap experiments. However, what if none of the genes contain buffered variation? The candidate region has recently been annotated with another gene, the non-coding RNA At2g22482 (Lamesch et al. 2012). Hsp90-dependent variation has been traced to a non-coding portion of the genome, a 3' UTR of a gene (Jarosz & Lindquist 2010); thus is it possible that the non-coding RNA harbors Hsp90 dependent variation. If none of the five genes tested are shown to harbor Hsp90 dependent variation in the allele swap experiments, this non-coding RNA is another candidate to consider before assuming that the Hsp90-buffered variation lies in intergenic regions.

If the region is indeed intergenic and outside of the non-coding regions included in the allele swap experiments, expression analyses of the five candidate may help uncover which variants are buffered by Hsp90. Expression of the five candidate genes would need to be completed in both the L*er* and CVI background. If expression of any of these genes differ upon

Hsp90 inhibition, the intergenic region nearby may harbor genetic variation that affects expression. Expression is known to be primarily affected by proximal promoter sequences and intronic sequence, so only in combination with the allele-swap experiments would more distant intergenic regions be implicated.

Methods

Plant materials and growth conditions

All lines were obtained from the ABRC: SAIL_297_F05, SALK_003443, SALK_025730, SAIL_401_C01, SALK_111466, SALK_054901 (At2g22480 SALK1), SALK_067305 (At2g22480 SALK2), SALK_090722 (At2g22480 SALK3). These lines were maintained through single seed descent and bulking to reduce the effects of spreading T-DNA insertion silencing. Hypocotyl assays were performed as described in Queitsch et al 2002. Ethanol sterilized seeds were planted on 1× Murashige and Skoog (MS) basal salt medium supplemented with 1× MS vitamins, 0.05% MES (wt/vol), 1% sucrose (wt/vol) and 0.3% (wt/vol) phytagel in a semi-randomized design with n = 70 per line.

Phylogeny

The amino acid sequences of the seven PFK were obtained from TAIR (Swarbreck et al. 2008). MUSCLE was used for amino acid alignment (Edgar 2004). Phylogeny was created using PhyML with default settings (Guindon et al. 2010).

Allele swaps

The candidate region spans five coding regions: At2g22460 (protein of unknown function, DUF617), At2g22470 (*ARABINOGALACTAN PROTEIN 2, AGP2*), At2g22475 (*GLABRA 2 EXPRESSION MODULATOR, GEM*), At2g22480 (*PHOSPHOFRUCTOKINASE 5, PFK5*), and At2g22490 (a D-type cyclin). For each of the five coding regions, we generated three *A. thaliana* transgenic lines in the respective coding region's T-DNA insertion background. Each T-DNA insertion line was transformed independently with the Col-0, CVI-0, and L*er*-1 ecotype allele of the respective coding region. Each ecotype allele was generated by PCR amplification using genomic DNA. Identical oligonucleotides were used to amplify the Col-0, Cvi-0, and Ler-1 alleles of each coding region. For four coding regions (At2g22460, At2g22470, At2g22475, At2g22480), PCR-generated alleles including the entire coding sequence as well as at least 1.0kbp of 5' and 3' regulatory DNA immediately upstream and downstream of the coding region. For At2g22490, the 3' regulatory region was truncated to approximately 0.1kbp for all three ecotypes.

All three alleles of all five coding regions were individually cloned into the pENTR1A plasmid (Invitrogen). We then used the Gateway LR Clonase II (Invitrogen) to recombine the alleles into a variant of the pK7WG2 plasmid (Karimi et al. 2002), from which we had previously excised the -35S promoter region. We then transformed each plasmid into Agrobacterium tumefaciens by electroporation and selected for successful transformed *A. tumefaciens* lines using kanamycin (50μg/mL). We then transgenically introduced each allele of each coding region into its respective T-DNA insertion line background using the floral dip method (Clough & Bent 1998). Transformants were selected on kanamycin (50μg/mL) and

propagated for at least three generations to generate lines that were homozygous for the transgenic insertion.

Main Figures

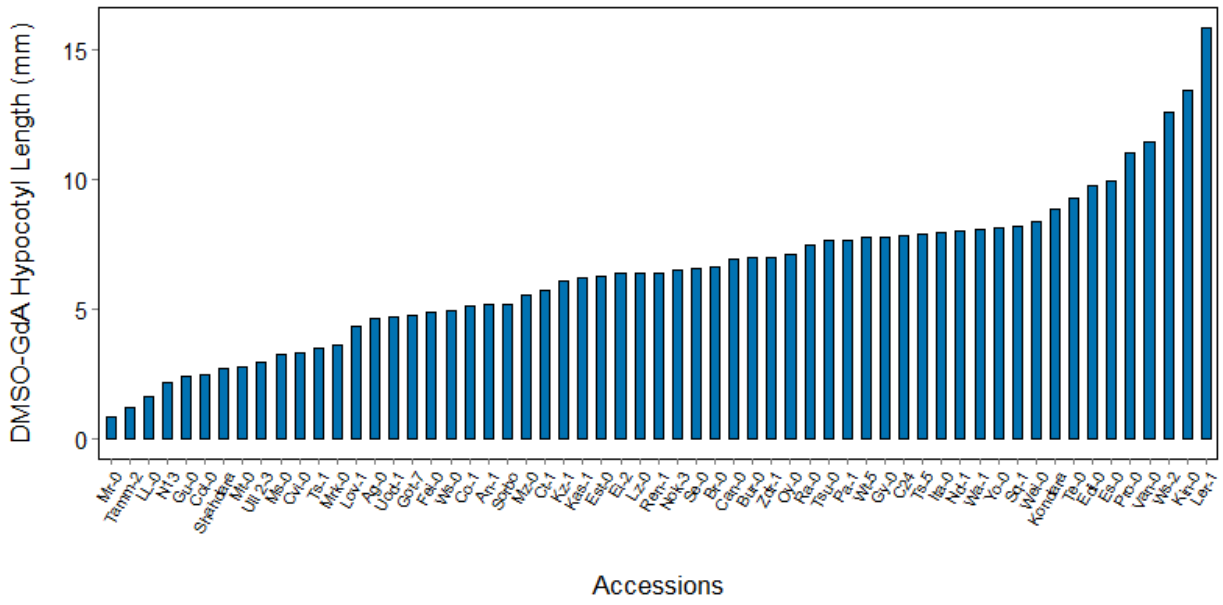

Figure 1. A wide range response to Hsp90 inhibition is measured across accessions for the trait dark grown hypocotyl. Response to Hsp90 inhibition was measured for the trait dark grown hypocotyls in 60 *A. thaliana* accessions. L*er* and CVI, the strains used for QTL mapping of the trait, lie in opposite extremities of the range.

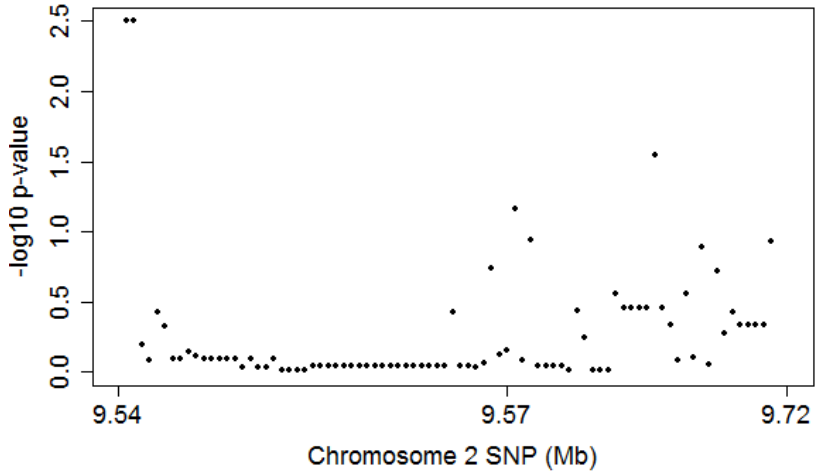

Figure 2. A candidate region containing Hsp90-buffered variation is identified using association mapping. Association was completed as described by Sangster (2007). The two SNPs with the greatest association ($p = 0.004$, Wilcoxon rank sum test) are 38bp apart. The uneven sampling across the region is due in part to the removal of variants with low minor allele frequency (<0.13).

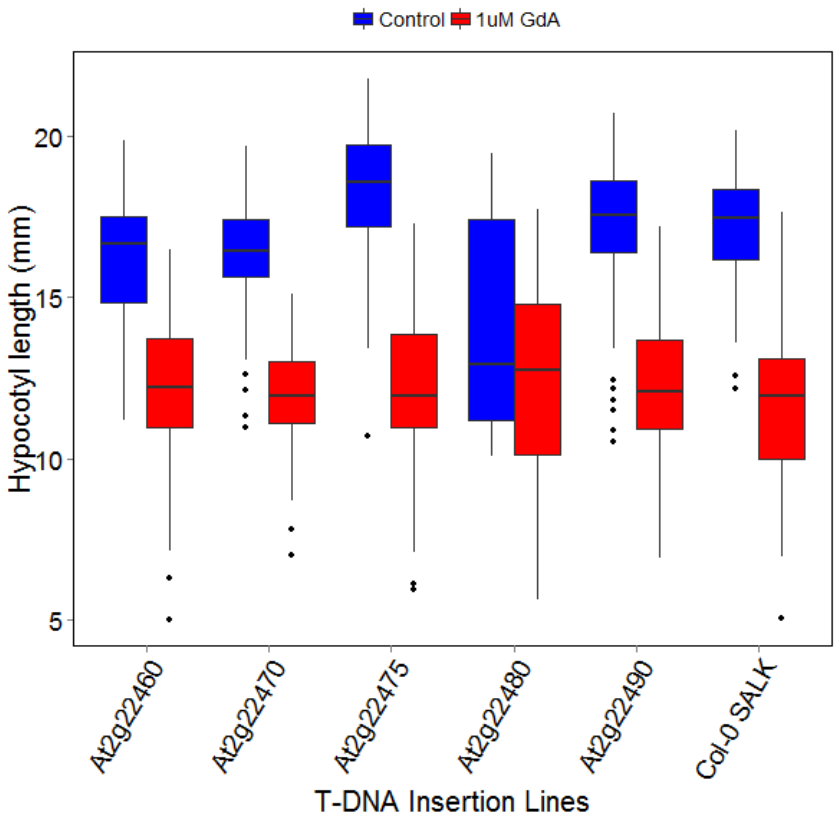

Figure 3. A candidate gene encoding Hsp90-buffered genetic variation is identified using T-DNA insertion lines. Seedlings from bulked seeds were grown for seven days in the dark on control media or media containing 1μM GdA. A T-DNA insertion in the gene At2g22480 exhibits lack of response to Hsp90 inhibition ($p = 0.15$, Tukey's post-hoc test). The other T-DNA insertion lines exhibit a response indistinguishable from wild-type (Col-0 SALK).

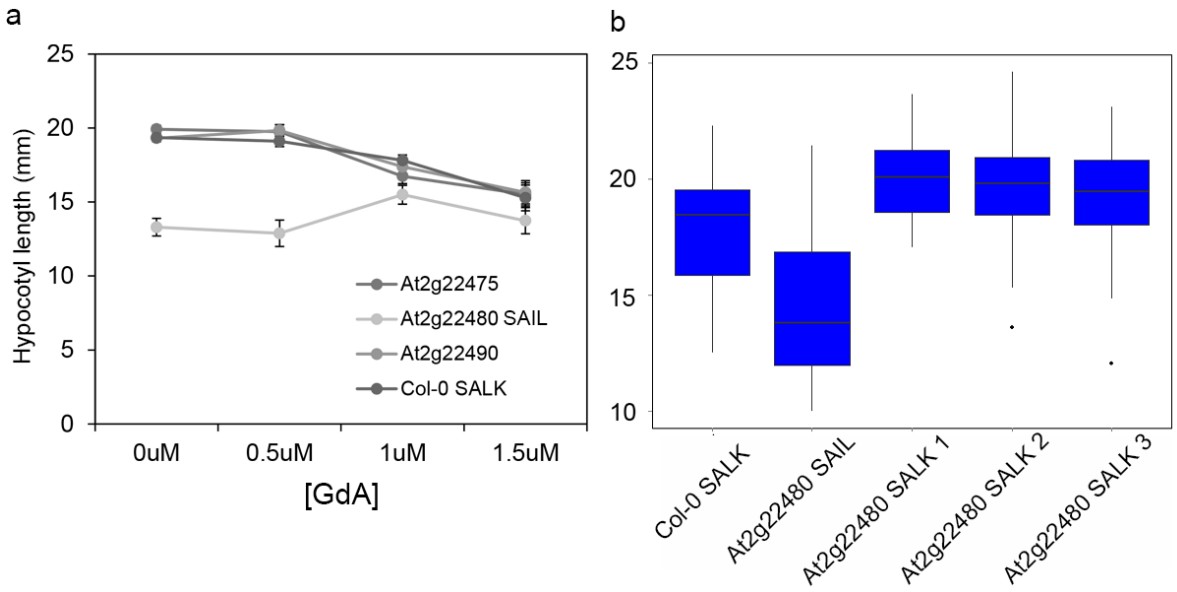

Figure 4. The At3g22480 SAIL T-DNA insertion line may be buffered by Hsp90. a) The SAIL T-DNA insertion in At2g22480 is not responsive to increasing degrees of Hsp90 inhibition. b) At2g22480 SALK1 and 2 T-DNA alleles are significantly taller than Col-0 SALK ($p < 0.01$, Tukey's post-hoc test) and the At2g22480 SAIL T-DNA line was significantly shorter ($p < 0.0001$, Tukey's post-hoc test). SALK3 was trending towards a increase in hypocotyl length ($p = 0.109$, Tukey's post-hoc test).

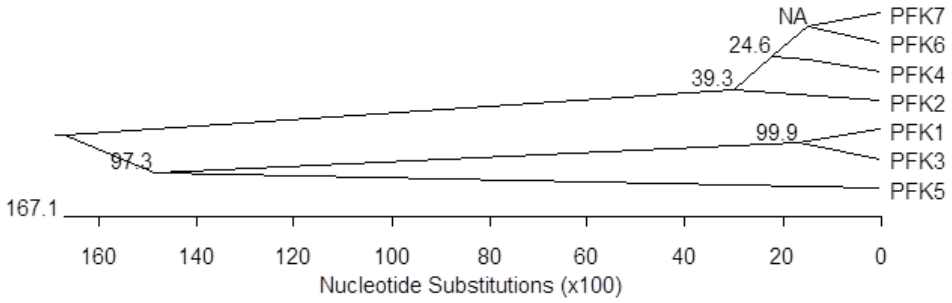

Figure 5. Phylogeny of *PFK* family in *A. thaliana*. The *PFK* family encodes seven genes in A. thaliana. *PFK5* is the basal branching member of the family (Mustroph et al 2007). Bootstrap support values are presented on the nodes.

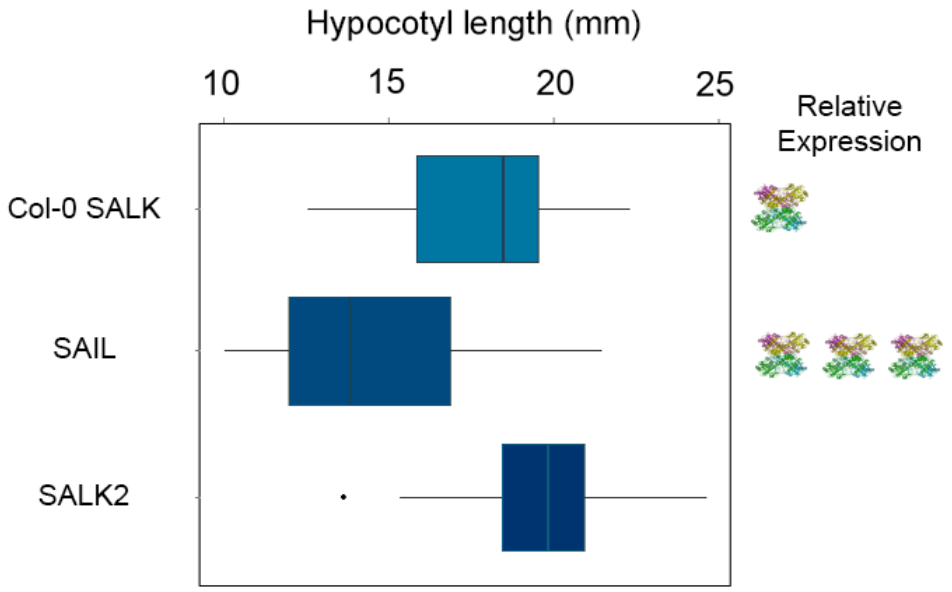

Figure 6. Model of *PFK5* T-DNA insertion line hypocotyl growth. *PFK5* is 3-fold overexpressed in the SAIL line compared to wild-type, as represented by the number of PFK molecules, and no expression is assumed in SALK2. The SAIL line is significantly shorter than wild-type, and the SALK2 line is significantly longer, which correlates to *PFK5* expression level (Figure 4b).

Supplemental figures

Ecotype	Stock Center Number		Ecotype	Stock Center Number
Ag-0	CS22630		Mz-0	CS22636
An-1	CS22626		N13	CS22621
Br-0	CS22628		Nd-1	CS22619
Bur-0	CS22656		Nok-3	CS22643
C24	CS22620		Oy-0	CS22658
Can-0	CS6660		Pa-1	CS6825
Col-0	CS22625		Pro-0	CS22649
Ct-1	CS22639		Ra-0	CS22632
Cvi-0	CS22614		Ren-1	CS22610
Edi-0	CS22657		Se-0	CS22646
Ei-2	CS22616		Shahdara	CS22652
Es-0	CS6699		Sorbo	CS22653
Est-0	CS22629		Sq-1	CS22600
Fei-0	CS22645		Tamm-2	CS22604
Got-7	CS22608		Te-0	CS6918
Gu-0	CS22617		Ts-1	CS22647
Gy-0	CS22631		Ts-5	CS22648
Kin-0	CS22654		Tsu-0	CS22641
Kondara	CS22651		Ull2-3	CS22587
Kz-1	CS22606		Uod-1	CS22612
L*er*-1	CS22618		Van-0	CS22627
LL-0	CS22650		Wa-1	CS22644
Löv-1	CS22574		Wei-0	CS22622
Lz-0	CS22615		Ws-0	CS22623
Mr-0	CS22640		Ws-2	CS22659
Mrk-0	CS22635		Wt-5	CS22637
Ms-0	CS22655		Yo-0	CS22624
Mt-0	CS22642		Zdr-1	CS22588

Table S1. Accessions analyzed for association mapping studies.

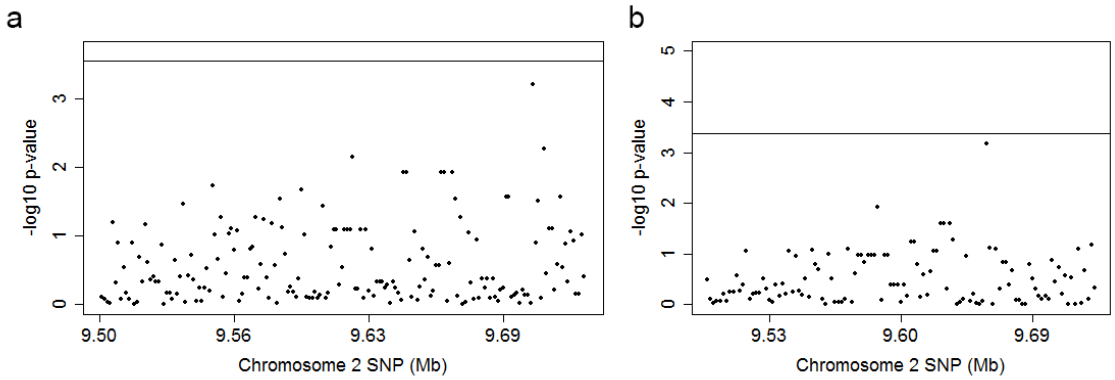

Figure S1. No associations are detected using dense genotyping data. a) Since Sangster (2007) was able to narrow the region of association to 30kb with Sanger sequencing every 3kb, we attempted to narrow the region further with publically available data in which accessions were genotyped every 500bp on average. No significant associations were detected using Wilcoxon rank sum test after Bonferroni correction (solid line). b) Sangster (2007) only sequenced the accessions with the most extreme responses to Hsp90 inhibition. Here the top and bottom quantiles of accessions were only included in the tests for association. Again, no significant associations were detected using Wilcoxon rank sum test after Bonferroni correction (solid line).

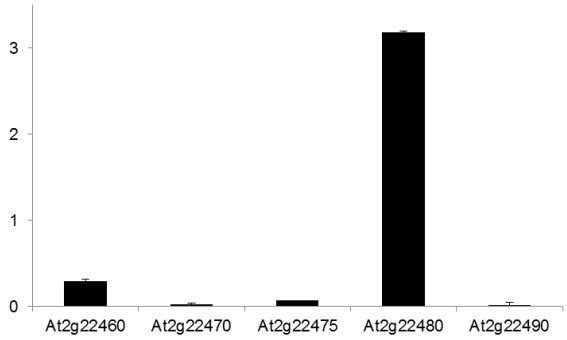

Figure S2. At2g22480 expression is not reduced in its respective T-DNA insertion line as is observed for the other candidate genes. At2g22480 appears to be more highly expressed in the SAIL T-DNA insertion line, possibly due to feedback regulation, though that remains to be tested.

CHAPTER V

DEGENERACY, FEEDBACK, AND ROBUSTNESS ACROSS THE *BZR*/*BEH* FAMILY[1]

Introduction

Identifying the functions of unstudied genes is both facilitated (Eisen 1998) and hindered by the presence of their paralogs. Paralogs can encode the same protein domains, which suggests that they function similarly. However, if the proteins are redundant or degenerate, mutations in one paralog may not have a phenotype, thus obstructing further studies. In *Arabidopsis thaliana*, there is an extensive presence of paralogs. About one-third of all *A. thaliana* genes make up one-thousand multi-member gene families (Swarbreck et al. 2008). Multiple whole genome duplications during angiosperm evolution are thought to have given rise to these gene families in part (Jiao et al. 2011; Blanc et al. 2003), and three well-supported whole genome duplications are detected in the *A. thaliana* lineage (Bowers et al. 2003; Simillion et al. 2002). In addition, *A. thaliana* exhibits a greater number of segmental and tandem duplication events compared to *Drosophila melanogaster* and *Caenorhabditis elegans* (Initiative 2000), contributing to both the large number and size of gene families present. Across these *A. thaliana* multi-member families, the function of the founding members of the gene family are often ascribed to the

[1] This chapter is in preparation for submission with co-authors Alex Mason, Anahit Galstyan, Karla Schultz, Jennifer Nemhauser, and Christine Queitsch to the *Journal of Experimental Botany*.

uncharacterized relatives (Eisen 1998), though evolutionary theory suggests that these family members should diverge in function (Zhang 2003).

Both empirical studies and population genetics theory find that the paralogs of gene families can diverge in a number of ways. After duplication and fixation in a population, most duplicated genes become pseudogenes and are lost from the genome (Lynch & Conery 2000), but others may be maintained, acquire new function (neofunctionalization), or maintain a subset of ancestral functions (subfunctionalization) (Conant & Wolfe 2008; Innan & Kondrashov 2010). For a gene to maintain function, the increased dosage of a gene cannot be harmful to fitness (Lynch & Conery 2000; De Smet et al. 2013). When paralogs do maintain the same function, they often exhibit diverged expression patterns (Force et al. 1999). The multiple copies can vary in expression by degree (Duarte et al. 2006), in time (Liu & Charng 2013), and/or by location (Liu et al. 2011). Comparisons of functions among family members can reveal the patterns that contributed to the maintenance of paralogs (Kramer et al. 2004).

When gene duplication occurs, the properties of genetic networks are altered (Conant & Wolfe 2008). Robustness is the genetic network property that is best-studied in regards to gene duplication (Gutiérrez & Maere 2014). Robustness is the ability of a system to withstand perturbations, such as mutation, environmental change, or developmental noise (Whitacre 2012). Gene duplications promote robustness by providing redundancy to genetic networks (Wagner 1996; Edelman & Gally 2001). For example single loss-of-function mutations commonly result in wild-type phenotypes, whereas mutations in two duplicated genes disrupt phenotypes (Wagner 2005; Edelman & Gally 2001). One proxy for robustness is variance in phenotypes among isogenic individuals growth in identical environments (Queitsch et al. 2002; Gibson & Dworkin

2004; Sangster et al. 2007; Todd A Sangster et al. 2008; Lempe et al. 2012; Folta et al. 2014). However, it is rare that the variance in a phenotype is examined in terms of the contribution of duplicated genes.

Here, we examine the *BEH* family of transcription factors in *A. thaliana* to explore their overlap in function and contribution to robustness. The founding members of the *BES1/BZR1 HOMOLOG* (*BEH*) family, *BRI1-EMS-SUPRESSOR1* (*BES1*) and *BRASSINAZOLE-RESISTANT1* (*BZR1*) were duplicated in the most recent whole genome duplication in the *A. thaliana* lineage (Blanc et al. 2003) and are the primary transcription factors in brassinosteroid signaling (Wang et al. 2002; Yin et al. 2002; Zhao et al. 2002). Brassinosteroid signaling regulates a large number of physiological processes in plants, ranging from seed maturation to senescence (Clouse 2002). Brassinosteroids are detected by the membrane-associated receptor BRI1 that then inhibits the activity of the GSK3 kinase BIN2. In the absence of brassinosteroids, BIN2 phosphorylates and inhibits BES1 and BZR1 (Zhao et al. 2002). In this phosphorylated state, BES1 and BZR1 are in association with the 14-3-3 proteins, which prohibit BES1 and BZR1 from entering the nucleus (Gampala et al. 2007). When brassinosteroids bind BRI1, BIN2 is inhibited, and the phosphatase PP2A de-phosphorylates BES1 and BZR1 (Tang et al. 2011) and causes their dissociation from the 14-3-3 proteins, allowing BES1 and BZR1 to localize to the nucleus. Once in the nucleus BES1 and BZR1 both activate and repress different sets of targets genes (He et al. 2005; Yin et al. 2005). Understanding the functional similarities and differences, not only between BES1 and BZR1, but also among all the BEH family members, will help clarify the regulation underlying many plant physiological processes.

BEH1-4 are little studied compared to *BES1* and *BZR1*. All four proteins share similarity to DNA-binding domain of BZR1 (Wang et al. 2002; He et al. 2005), therefore likely acting as transcription factors as well. BEH1-4 are phosphorylated in a manner similar to BES1 and BZR1 (Yin et al. 2005), and yeast two-hybrid analyses have shown that BEH2, in addition to BES1 and BZR1, is phosphorylated by another GSK3 kinase, ASKϴ (Rozhon et al. 2010). Moreover, microarray studies show that the family members are differentially expressed throughout development (Hruz et al. 2008). Here we systematically examined the entire *BEH* family for phenotypes. We found partial redundancy in function among members and identified a new target gene for studies in auxin-brassinosteroid crosstalk. We identified environmentally-dependent genetic interactions between family members. Specifically one member of the family was critical for robustness in a brassinosteroid-related trait. Combining these findings, we propose a model of *BEH* family function distribution and feedback regulation.

Results

T-DNA insertions in the BEH *family do not affect examined adult plant traits*

The *BEH* family encodes six members, including *BEH1-4*, for which very little functional information is available (Yin et al. 2002; Wang et al. 2002; Yin et al. 2005; Rozhon et al. 2010). *BES1* and *BZR1* were discovered in screens for plants could suppress brassinosteroid insensitivity (Yin et al. 2002) or suppress an inhibition of brassinosteroid synthesis (Wang et al. 2002), respectively. The same amino acid change was identified in both dominant *bes1-D* and *bzr1-1D* in these screens, which is positioned in the PEST domain of these two proteins. This homologous domain is not conserved among all the other family members (Figure S1) (Rogers et

al. 1986), preventing the creation of similar dominant mutations in all family members. We instead acquired T-DNA insertion mutants of each member of the family (Figure S1) to uncover the functions of the family members (Lamesch et al. 2012). The respective gene products were not detected in the bacterial transferred DNA (T-DNA) insertion lines.

We examined mutants in each family member for adult phenotypes using these T-DNA lines. Flowering time regulation is regulated by brassinosteroids (Jihong Li et al. 2010), so we suspected that the *beh* family mutants would exhibit altered flowering time. For example, BES1 interacts with the flowering time regulating proteins, ELF6 and REF6 (Yu et al. 2008). We noted no difference in flowering time among the mutants (Figure S2, $n > 15$, 1 rep). Of these six mutants, two were also examined by Galstyan (pers. comm), who also observed no adult phenotypes for *bes1-2* and *beh4-1*. Supporting our observations, Yin et al (2005) detected no obvious adult phenotypes with a *BES1* T-DNA insertion line. Overall, we conclude that adult *BEH* family mutant plants exhibited no obvious phenotypes.

The basal members of the family promote growth the most of the family in an environmentally-dependent manner

As there were no detectable phenotypes in adult traits among the mutants, we focused on seedling phenotypes. The genotypes *bes1-2, bzr1-2, beh3-1,* and *beh4-1* hypocotyls grew significantly less than wild-type controls in the dark (Figure 1, $p < 0.0001$, linear mixed effects model, $n = 70$), indicating that these four factors are positive regulators of dark growth. These results are consistent with previous findings that RNAi targeting *BES1* reduces hypocotyl length (Yin et al. 2005; Wang et al. 2013), and that the *bes1-1* T-DNA insertion line exhibits reduced

hypocotyl length (He et al. 2005). Of the four mutants different from wild-type, *beh4-1* exhibited the greatest growth defect (Figure 1). We conclude that knock-outs in the best-studied and founding members of the *BEH* family, *BES1* and *BZR1*, were not the most affected in dark growth; however, knock-outs of the most ancestral members of the BEH family, *BEH3* and *BEH4* had the greatest effect on dark growth.

When grown in the light, *bes1-D* and *bzr1-1D* exhibit opposing effects on cell elongation, with *bzr1-1D* exhibiting a shortened hypocotyl (He et al. 2005; Gampala et al. 2007). Previously, reduced growth was observed in the light in the T-DNA insertion line, *bes1-1* (He et al. 2005); therefore, we examined T-DNA mutants across the family for light growth. We hypothesized that the *bzr1-2* would be longer than wild-type in the light, based on the *bzr1-1D* phenotype. Because light grown seedlings are short in length, we completed a power analysis to determine the number of seedlings we should measure. To detect an effect size of 0.5mm at a power of 0.8, we required 70 seedlings. In one of the three replicates, *bzr1-2* was significantly longer than wild-type, but this was not replicable in the additional experiments.

Overall, there were no significant differences between any of the genotypes and their respective wild-type seedlings in the light (Figure 1b). These results suggested several possibilities. First, the effect size differences we could detect with our experimental design may be larger than the effect size that the mutants have on the phenotype. Second, the result with *bes1-1* was observed in different light conditions (He et al. 2005). Third, the mutants may act redundantly in the light in contrast to the dark.

Environmentally-dependent epistatic, additive, and antagonistic interactions detected among family members

Many studies of BR signaling only examine BES1 or BZR1, perhaps assuming that *BES1* and *BZR1* act redundantly (Yu et al. 2008; Li et al. 2009, 2010; Jin et al. 2013). Although both *bes1-D* and *bzr1-D* do exhibit increased hypocotyl growth in the dark, and *BES1* and *BZR1* share 88% amino acid identity, evidence abounds in the literature that the proteins differ. Comparisons of ChIP-chip (Yu et al. 2011; Sun et al. 2010) experiments have found that these two proteins have many independent targets, and the preferred promoter binding motifs of BES1 and BZR1 are different (He et al. 2005; Yin et al. 2005). Specifically, *BZR1* acts in negative feedback control of BR synthesis, whereas BES1's contribution to BR feedback is lower (Wang et al. 2002; Mathur et al. 1998). Our results in the dark indicated that *BES1* and *BZR1* share a function in promoting hypocotyl growth, but do not act redundantly based on significant differences in hypocotyl length in both single mutants compared to wild-type (Figure 1).

To understand more about the similarity of function between *BES1* and *BZR1*, we compared the effects of the single mutants to each other in the dark. We found no significant difference in dark growth between *bes1-2* and *bzr1-2*, suggesting that *BES1* and *BZR1* contribute to dark growth to the same degree (Figure 2a). To determine whether BES1 and BZR1 independently regulate dark growth, we examined *bes1-2;bzr1-2*. The double mutant tended to be shorter than either single mutant, but the double mutant was only significantly shorter than *bes1-2*, indicating that *BZR1* is epistatic to *BES1* in the promotion dark hypocotyl growth. This result is in contrast to the finding that the *bes1-2;bzr1-1* does not differ from wild-type in the light (Galstyan pers. comm). Therefore, light conditions regulated the interaction of these family

members. Such environmental dependency of genetic interaction is a frequent occurrence (Koornneef et al. 1998; Harrison et al. 2007) and supports that a complicated interplay exists among the *BEH* family members. Notably, the *beh4-1* single mutant was still shorter than the *bes1-2;bzr1-2* double mutant, demonstrating a large role for *BEH4* in dark growth.

Galstyan (pers. comm.) observed that *bes1-2;beh4-1* exhibited an additive phenotype compared to the single mutants when growth in the light and in response to BL. Both *bes1-2* and *beh4-1* exhibited a reduction in hypocotyl growth in the dark (Figure 1) indicating that the two genes are not redundant in regulating dark growth. We examined whether *BES1* and *BEH4* also act additively in the dark. The *bes1-2;beh4-1* double mutant revealed that *beh4-1* phenotype was partially rescued by the addition of the *bes1-2* mutation, demonstrating that the two genes acted antagonistically in the dark. Therefore, light conditions again altered the interaction of *BEH* family members.

Among the family members, *BEH1* and *BEH2* did not affect dark growth (Figure 1). The suggested redundancy among the family members spurred us to test whether *BEH1* and *BEH2* were redundant in the regulation of dark growth. Neither *beh1-1;beh2-1* nor the single mutants exhibited a growth difference compared to wild-type (Figure 2c). These results suggest either that *BEH1* and *BEH2* do not regulate dark growth or that *BEH1* or *BEH2* are further redundant with other genes regulating dark growth.

In summary, we observed several patterns of interaction among the family members and extensive degeneracy in function. *BES1* and *BEH4* displayed a plastic interaction, switching from an additive interaction in the light to an antagonistic interaction in the dark. The most

closely related pair of genes in the family, *BES1* and *BZR1*, interacted epistatically in the dark, but the type of interaction between *BES1* and *BZR1* in the light cannot bet determined because no phenotype was observed. The same inability to detect interaction between *BEH1* and *BEH2* occurred in the dark.

Time course experiments reveal specific functions of BEH1, BEH2, *and* BEH3

We examined *beh1-1*, *beh2-1*, as well as *beh3-1* over a time course because no phenotypes were associated with *BEH1* and *BEH2* above (Figure 3) (Stewart et al. 2011). A time course allows for a close dissection of the functions of these transcription factors and may reveal phenotypes that cannot be observed when only considering endpoints. Previously *bes1-2*, *bzr1-2*, and *beh4-1* have been analyzed in the same experimental set-up, but no growth defects in the light were observed. In contrast, we observed several differences in light growth in the *beh1-1*, *beh2-1*, and *beh3-1* mutants. At days 5-6 *beh1-1* was significantly shorter than wild-type plants, but recovered and reached wild-type length at day 7 (Figure 3a). *beh2-1* exhibited a similar pattern, with significantly shorter hypocotyls at day 5, but *beh2-1* recovered one day faster, at day 6 (Figure 3a). In contrast to *beh1-1* and *beh2-1*, *beh3-1* exhibited an earlier growth defect, at day 4 and recovered by day 5 (Figure 3b). Indeed, *BEH1*, *BEH2*, and *BEH3* positively regulated hypocotyl growth in the light but at specific stages of hypocotyl elongation.

Regulation of hormone responses by BEH1, BEH2, *and* BEH3

BES1 and BZR1 are known to regulate the response to brassinosteroids. Indeed, Galstyan (pers. comm.) found that *bes1-2*, *bzr1-2* and *beh4-1* were hyposensitive to brassinosteroids (0.5μM brassinolide, BL) compared to controls; therefore, we tested the rest of the family for

response to BL. With BL treatment *beh1-1* no longer had a growth deficit at days 5-6 in the presence of BL (Figure 3c) compared to mock treatment (Figure 3a). In contrast, *beh2-1* was hypo-sensitive during day 4, but reached wild-type length, at day 5 (Figure 3c), indicating that the presence of BL rescued the growth deficit observed under mock treatment at day 5. During days 3 and 4, *beh3-1*, was also hypo-sensitive but reached wild-type growth on day 5 (Figure 3d), similar to reaching wild-type length under mock treatment as well. Overall *beh2-1* and *beh3-1* exhibited insensitivity to BL at early days and BL helped rescue hypocotyl lengths in *beh1-1* and *beh2-1* sooner compared to mock treatment. These results indicate that both *BEH2* and *BEH3* were involved in BL response but acted during different times during seedling growth, and BL may act independently of *BEH1* in hypocotyl length regulation. *BEH3* gene expression is not known to be regulated by BL, but both *BEH1* and *BEH2* expression is down-regulated by BL (Table 1) (Sun et al. 2010), indicating complex regulation of growth promotion.

Many studies have revealed that BES1 and BZR1 are involved in cross-talk among different hormones, including, but not limited to, auxins (Clouse et al. 1992; Nemhauser 2008). Auxin (5μM picloram, PIC) rescued *beh1-1* and *beh2-1* phenotypes so that they were indistinguishable from wild-type (Figure 3e). With auxin treatment, *beh3-1* was hypo-sensitive compared to wild-type from day 3 and remained significantly shorter through day 8 (Figure 3f). Among these three family members, only *BEH3* was needed for auxin responses during seedling growth at all stages, and auxins may act independently of *BEH1* and *BEH2*.

Cross-talk between brassinosteroids and auxin in plant morphology regulation is well-established (Clouse et al. 1993; Nemhauser et al. 2004; Mouchel et al. 2006; Vert & Chory 2006). *BES1*, *BZR1*, and *BEH4* are known to bind the promoters of both auxin-regulated and

brassinosteroid-regulated genes (Galstyan pers. comm.). *beh1-1* growth phenotypes were rescued by the addition of PIC and BL, at the same developmental time that either single hormone rescued (Figure 3c, 3e, 3g). With PIC and BL treatment, *beh2-1* seedlings did not show growth deficits at earlier time-points, days 4-5, as observed in mock and BL treated plants. Rather *beh2-1* was hyper-sensitive at days 6-8 (Figure 3g), elongating to a greater degree than wild-type in the present of both BL and PIC. These results suggest that *BEH2* acted to inhibit auxin's growth promotion effects in the presence of BL at later time-points. With BL and PIC treatment, *beh3-1* were hypo-sensitive at days 3-4 (Figure 3h), which was identical to its response to BL alone. Addition of PIC, even in combination with BL, did not fully rescue the phenotype.

Expression feedback among members of the BEH family in the light and dark

The above analyses and previous studies (Yin et al. 2005) support the idea that across the *BEH* family, members are partially but not fully redundant, which can indicate feedback control among family members. BES1 and BZR1 ChIP-chip results (Sun et al. 2010; Yu et al. 2011) identified the regions upstream of every other family members as potential transcriptional targets of BES1 and BZR1 (Table 1). Additionally, expression of *BEH2* was up-regulated in RNAi lines in which *BES1* was targeted (Wang et al. 2013). We will explore the relative expression of each of the *BEH* family members in each of the single knock-out mutants. The different family members also act differently in varying light conditions, so we will complete this gene expression analysis in in both the light and dark.

Robustness in dark conditions is regulated by BEH4

Within the *BEH* family, we noticed that *beh4-1* dark grown hypocotyls had a larger variance than the other mutants (Figure 1). This increased variance indicated reduced robustness to developmental noise as the individuals were isogenic and reared in the same environment (Waddington 1942; Queitsch et al. 2002). Since redundancy is one feature of genetic networks that provides robustness (Gu et al. 2003; Wagner 2000), we wondered whether the partially redundant members of the *BEH* family provide robustness to phenotypes they regulate. We calculated the coefficient of variance (CV, standard deviation/mean) for dark grown hypocotyls for each of the single mutants (Figure 4a), which is a common statistic to measure robustness (Lempe et al. 2012; Gutiérrez & Maere 2014). The CV of *beh4-1* was significantly greater than wild-type ($p = 0.03$, Welch's t-test, $n = 3$). Robustness in dark grown hypocotyls arose through the activity of *BEH4*, the member of the family that influences trait mean the most.

We hypothesized that a further loss of robustness would be observed by removing the function of additional members of the family. We examined the *bes1-2;beh4-1* mutant because both single mutants have an effect on mean hypocotyl length in the dark. Surprisingly, we found that the additional loss of BES1 activity partially rescued the loss of robustness observed in *beh4-1* (Figure 4b). This result is reminiscent of the rescue observed for hypocotyl mean in the *bes1-2;beh4-1*. BES1 and BEH4 appear to also have opposing functions in robustness in dark grown hypocotyls.

Hsp90's maintenance of robustness may be via BEH4

The best studied regulator of robustness in *A. thaliana* is Hsp90 (Queitsch et al. 2002; Sangster et al. 2007; Todd A Sangster et al. 2008; T Sangster et al. 2008). Hsp90 is a protein

chaperone that facilitates the proper folding of its substrates (Taipale et al. 2010). In hypocotyl length in *A. thaliana*, Hsp90 activity is necessary for wild-type levels of robustness to developmental noise (Todd A Sangster et al. 2008). As both *Hsp90* and *BEH4* have roles in hypocotyl robustness, we tested for their genetic interaction in robustness. We grew *beh4-1*, *bes1-2*, and wild-type seedlings in the presence of a potent and specific inhibitor of Hsp90, geldanamycin (GdA) and assessed the robustness of each of the genotypes using CV. Upon Hsp90 inhibition, we observed an increase in CV in both wild-type and *bes1-2*, whereas *beh4-1* did not exhibit a further increase in CV (Figure 5a). In fact, the *beh4-1* CV was not significantly different between the two treatments ($p=0.93$, Welch's t-test, $n=3$), demonstrating an interaction between *BEH4* and Hsp90 for hypocotyl CV. We conclude that the increase in variance due to HSP90 inhibition is via the loss of the activity BEH4 in these conditions.

Among the members of the BEH family, BES1, but not BZR1, has been shown to be an Hsp90 client (Lachowiec et al. 2013; Shigeta et al. 2013). It is possible that other family members are also Hsp90 substrates in addition to BES1, as client status is often shared among family members (Taipale et al. 2012). When Hsp90 is inhibited, the function of its substrates is compromised or altered (Taipale et al 2010). The lack in increase in CV upon Hsp90 inhibition is consistent with our hypothesis that BEH4 is an Hsp90 client. We analyzed the response of each of the single mutants of the *BEH* family members to GdA to explore whether members may be Hsp90 substrates. *bes1-2*, *beh3-1*, and *beh4-1* were less sensitive to HSP90 inhibition compared to wild type (Figure 5b). Since BES1 is a known Hsp90 substrate and BZR1 is not, these results are consistent with the hypothesis that BEH3 and BEH4 are also Hsp90 substrates.

Discussion

Functional annotation of *A. thaliana* genes is an ongoing process. Many genes have been categorized into gene families, but the individual genes have not been examined. Here we examine the rest of the family members of the *BEH* family, the family that encodes the primary transcription factors of brassinosteroid signaling. One trait that has been thoroughly examined for brassinosteroid phenotypes is dark grown hypocotyls (Chory 1999). Although *BES1* and *BZR1*, the founding members of the family, were identified for mutant hypocotyl phenotypes, we found that *BEH4* and *BEH3* have greater contributions in regulating hypocotyl length. *BEH4* was not only the greatest regulator of dark growth, but also maintained a stable, repeatable phenotype. None of the other single mutants exhibited an increase in variance. *BEH4*, which is the earliest branching member of the family, contributes the most in regulating the dark growth of hypocotyls.

Environment-dependent changes in the interactions among genes are not uncommon (Koornneef et al. 1998; Harrison et al. 2007). Galstyan (pers. comm.) observed that both *BES1* and *BEH4* act additively in regulating hypocotyl growth in the light. In the dark, we found that *BES1* and *BEH4* act antagonistically, indicating that there is an environmentally controlled aspect of their interaction. This antagonism was not only observed in the hypocotyl mean, but also in the hypocotyl variance. Based on these data, we conclude that family members can change between positive to negative regulation of growth depending on light conditions.

BES1 and *BZR1* regulate the transcriptional response to brassinosteroids (Zhiponova et al. 2014) and are important for hormone crosstalk (Clouse et al. 1992; Nemhauser et al. 2004; Mouchel et al. 2006; Vert & Chory 2006; Walcher & Nemhauser 2012). *BEH4* is also necessary for brassinosteroid and auxin responses (Galstyan pers. comm.). Here we examined *BEH1*,

BEH2, and *BEH3* for their roles in response to brassinosteroids and auxin. We observed that *BEH1* was not necessary for brassinosteroid or auxin responses. *BEH2* was partially necessary for early responses to brassinosteroids and also negatively regulated the response to brassinosteroids in combination with auxin at later stages. Brassinosteroid signaling during the first days of seedling growth required *BEH3*. The most striking finding was that *BEH3* was partially required for response to auxin at all time-points. Further studies of auxin-BL crosstalk should target *BEH3* as an important component. BZR1 and BES1 also integrate the inputs of brassinosteroids and gibberellic acid (GA) (Gallego-Bartolomé et al. 2012; Li et al. 2012; Oh et al. 2012; Bai et al. 2012; Jaillais & Vert 2012), though we did not detect differences in responses to GA among the family members (Figure S3). Several *BEH* family members may have roles in strigolactone signaling as well (Wang et al. 2013). The six members of the *BEH* family are degenerate in their roles in a number of hormone signaling pathways.

Close examination of the light grown time-lapse results in this study and data from Galstyan (pers. comm.) reveals which family members share function and at what time during development (Figure 6). *BEH3* was the only family member that regulated early light growth regardless of hormone treatment. *BZR1* was important for responses to brassinosteroid and auxin at later time points but additional genes always regulated those traits as well. Interestingly, *BES1* only acted to regulate phenotypes that are also regulated by *BEH4*, though the converse was not true. *BEH2* positively regulated early growth, but negatively regulated later growth in response to a combination of brassinosteroid and auxin. For the traits examined, *BEH1*, *BEH2*, and *BEH3* were the members of the family that regulated responses individually and *BES1*, *BZR1*, and *BEH4* had degenerate functions.

Although examination of all members of the family enabled us to detect some functional differences, many further questions remain. First, BES1 is known to form heterodimers with several other transcription factors (Yin et al. 2005); however, it is unknown whether heterodimers form between *BEH* family members. Second, the patterns of gene expression of the family members across cell types and development are unknown. Third, there is no information about the promoters bound by BEH1-4. Combining these three pieces of information would further reveal how function has been subdivided among *BEH* family members.

Methods

Plant materials and growth conditions

bes1-2 (Lachowiec et al. 2013), *bzr1-2* (GABI_857E04), *beh3-1* (SALK_017577), and *beh4-1* (SAIL_750_F08) are in the Col-0 background. *beh1-1* (SAIL_40_D04) and *beh2-1* (SAIL_76_B06) are in the Col-3 background. For hypocotyl length assays, seeds were sterilized for 10 minutes in 70% ethanol, 0.01% Triton X-100, followed by 5 minutes of 95% ethanol. After sterilization, seeds were suspended in 0.1% agarose and spotted on plates containing 0.5x Murashige Minimal Organics Medium and 0.8% bactoagar. Seeds on plates were then stratified in the dark at 4°C for 3 days and then transferred to an incubator cycling between 22° for 16 hours and 20° for 8 hours to imitate long days. Plate position was changed every 24 h to minimize position effect for light grown seedlings. Racks of plates containing dark-grown seedlings were wrapped in foil. For hormone or inhibitor assays, the additives were suspended in the media in the following concentrations: 1μM geldanamycin (Sigma), 0.5μM brassinolide (), 5μM picloram, and 5μM GA3. Equivalent amounts of solvent were used for control treatment.

As solvents DMSO was used for geldanamycin, brassinolide, and GA, and 80% ethanol was used for picloram.

Phenotyping

For estimates of hypocotyl CV, three replicates of n > 50 were measured. Assays of mean hypocotyl length were completed in triplicate with n > 15. Photos were taken of each plate, and individual hypocotyls were manually measured using NIH ImageJ1.46r. Days to flowering, as measured by the appearance of the flowering meristem, was examined for each genotype (n > 10).

qPCR

Two biological replicates of sixty pooled 5-day light-grown seedlings and three biological replicates of sixty pooled 5-day dark grown seedlings were harvested. Tissue was frozen in liquid nitrogen and ground by hand with a pestle. RNA was extracted using the SV Total RNA Isolation kit (Promega). To remove contaminating DNA, a second DNase treatment was completed according to the Turbo DNase protocol (Ambion). Poly-A tail cDNA was produced using LightCycler kit with oligo-dT primers (Life Technologies). qPCR primers are listed in Table S1. In both *bzr1-2* and *beh2-1* these qPCR primers amplified products. The absence of the full length transcript was confirmed using primers that target the full-length transcript.

Time-lapse hypocotyl assay

Time-lapse photography was as previously described (Stewart et al. 2011). Briefly, images were taken every 24 hours with a charge-coupled device camera (PL-B781F, PixeLINK, http://www.pixelink.com/) equipped with a lens (NMV-25M1, Navitar, http://www.navitar.com/) and an infrared long-pass filter (LP830-35.5, Midwest Optical Systems, Inc., http://www.midopt.com/). A custom LabVIEW (National Instruments, http://uk.ni.com/) program controlled image capture and illumination. Time-lapse assays were completed in short-day conditions.

Main Figures

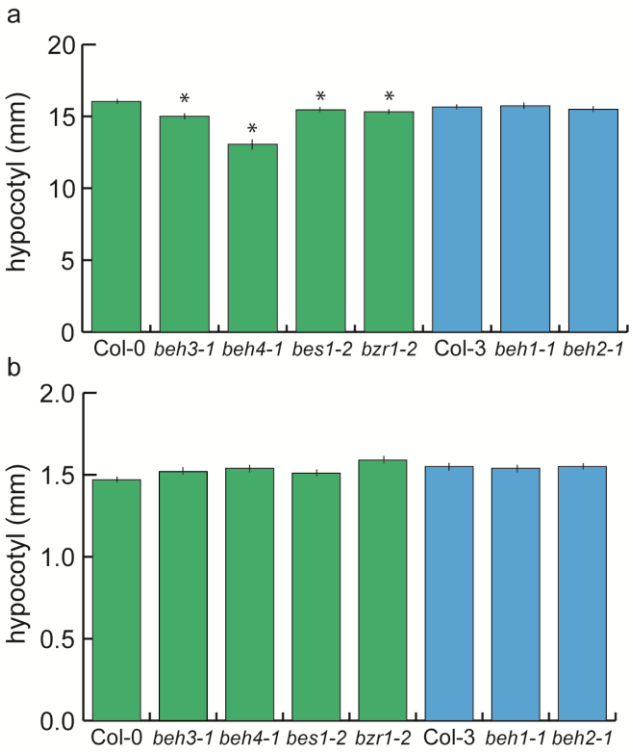

Figure 1. The *BEH* family encodes genes that singly regulate hypocotyl growth in the dark, but not the light. a) Seedlings were grown for seven days in the dark and hypocotyls were measured. *beh3-1*, *beh4-1*, *bes1-2*, *bzr1-2* grew significantly less than wild-type ($p < 0.0001$, linear mixed effects model with genotype as a fixed effect and replicate as a random effect). b) Mutant seedlings grown in long-day conditions did not exhibit altered growth compared to wild-type. One representative replicate with standard error of the mean for n = 70 is shown.

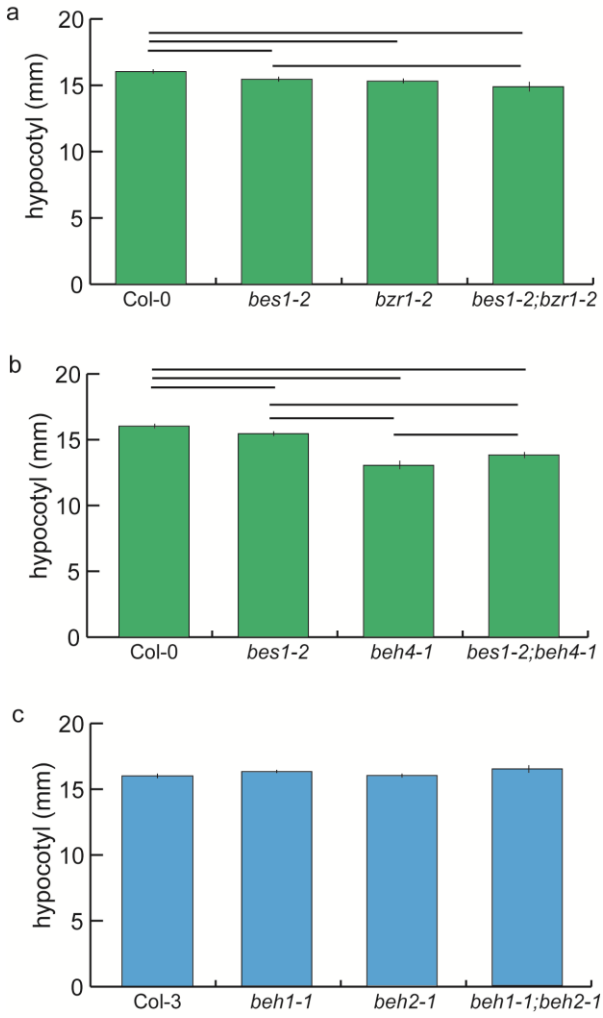

Figure 2. Double mutants in the *BEH* family exhibit multiple types of genetic interactions.

a) Examination of the *bes1-2;bzr1-2* mutant indicates that *BZR1* is epistatic to *BES1* in dark growth as observed in the lack of significant difference in growth between *bes1-2* and *bes1-2;bzr1-2*. Significant differences ($p < 0.05$) are displayed by the horizontal bars as determined by linear mixed effect modeling. b) *BES1* and *BEH4* interact antagonistically as demonstrated by the intermediate phenotype of their double mutant. c) No differences in dark growth are observed for *beh1-1*, *beh2-1*, or for *beh1-1;beh2-1*. For a-c) one representative replicate with standard error of the mean for $n > 20$ is shown.

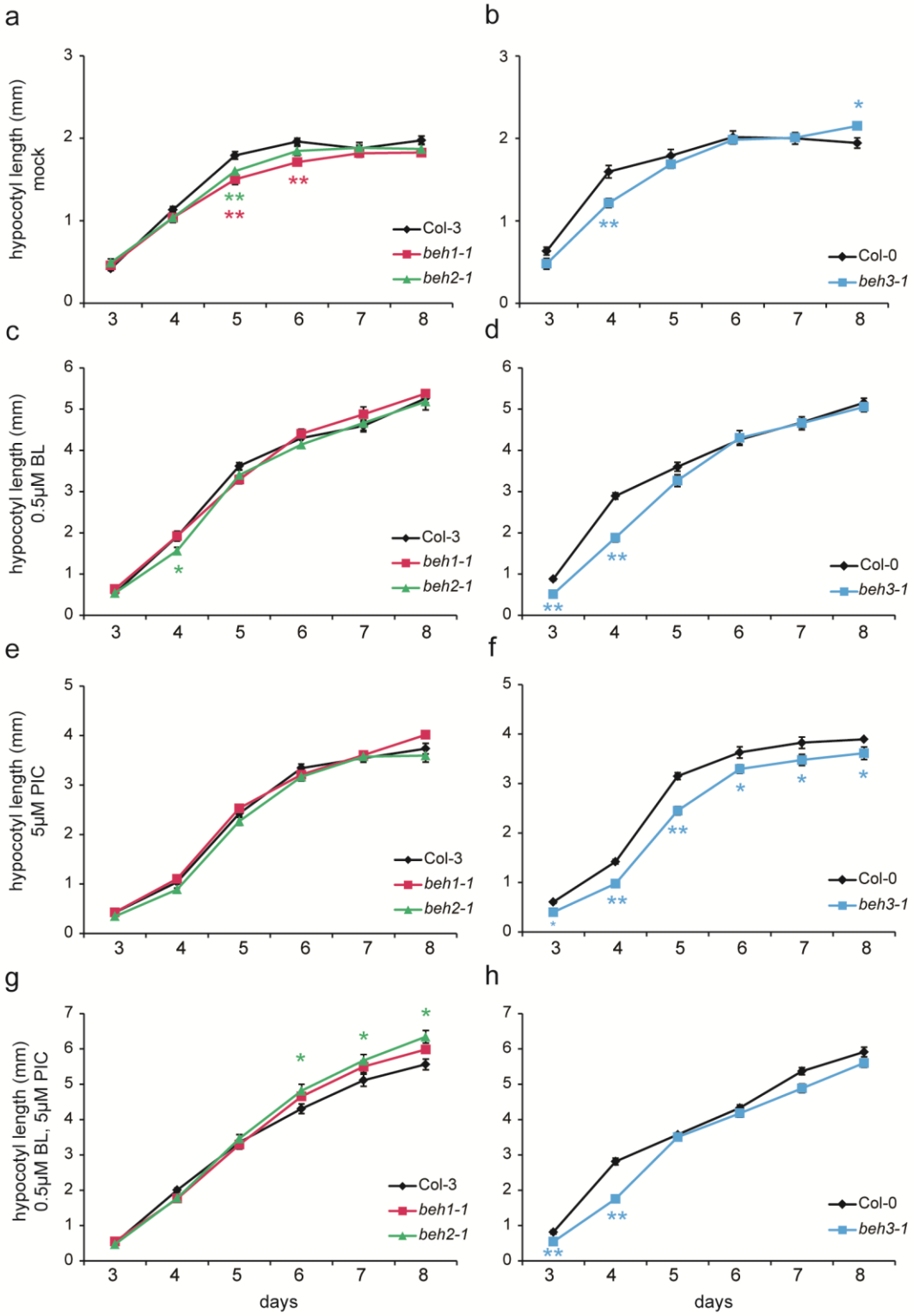
a
b
c
d
e
f
g
h
hypocotyl length (mm)
mock
0.5μM BL
5μM PIC
0.5μM BL, 5μM PIC
days
Col-3
beh1-1
beh2-1
Col-0
beh3-1

Figure 3. Functions of *BEH1* and *BEH2* are dependent on developmental stage. a,b) Growth of mock treated mutants and corresponding wild-type were tracked every 24 hours from days 3-8 post-germination, revealing subtle defects in growth rate. c,d) Growth rate tracking with brassinosteroid treatment exposes early deficits in BL response. e,f) *BEH3*, but not *BEH1* or *BEH2,* is needed for proper response to auxins. g,h) A surprising role for *BEH2* in repressing a brassinosteroid/auxin combinatorial response was discovered. Significance was determined with Welch's t-test: *$p < 0.05$, **$p < 0.01$.

Gene	Up-regulated with BR[1]	Down-regulated with BR[1]	Target of BES1[2]	Target of BZR1[1]
At1g19350 BES1	X		X	X
At1g75080 BZR1	X			X
At3g50750 BEH1		X	X	X
At4g36780 BEH2		X	X	X
At4g18890 BEH3				X
At1g78700 BEH4			X	

Table 1. Co-regulation of family members and response to brassinosteroids.

[1] Sun et al 2010, [2]Yu et al 2011. Differences in BES1 and BZR1 targets may be due to differences in the ChIP assays. Targets of BES1 were determined with ChIP-chip using an anti-BES1 antibody on material collected from *bes1-D* 14d seedlings on plates or the BZR1 ChIP-chip (Yu et al. 2011).Targets of BZR1 were determined with ChIP-chip using an anti-YFP antibody on rosette tissue collected from BZR1-CFP plants grown at LD for 4 weeks on soil (Sun et al. 2010).

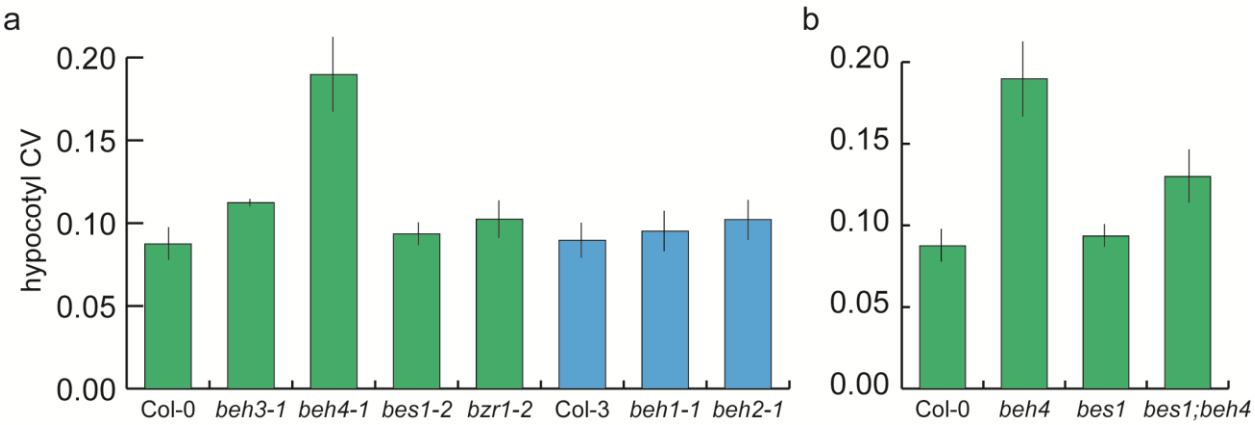

Figure 4. *BEH4* regulates robustness to developmental noise in dark grown hypocotyls. a) *BEH4* exhibits significantly greater variation in hypocotyl length compared to wild-type (p = 0.03, Welch's t-test). The CV was estimated in three biological replicates. Standard error of the mean for n = 3 is shown. b) The double mutant *bes1-2;beh4-1* exhibits a variance intermediate to either single mutant.

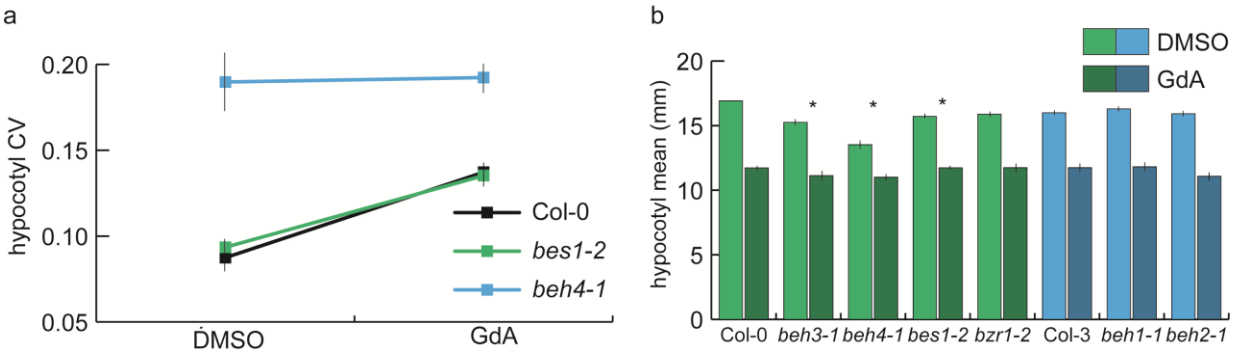

Figure 5. Dark grown hypocotyl robustness provided by Hsp90 arises from its interaction with *BEH4*. a) Hypocotyls were grown on mock (DMSO) and Hsp90-inhibited (GdA) media in three biological replicates. The CV was calculated for each replicate and the standard error of the mean for n = 3 is shown. b) Seedlings were grown on mock (DMSO) and Hsp90-inhibited (GdA) media. *Significant differences in response to Hsp90 inhibition are shown ($p < 0.03$, linear mixed model with genotype, treatment, and interaction effects as fixed effects and replicate as a random effect).

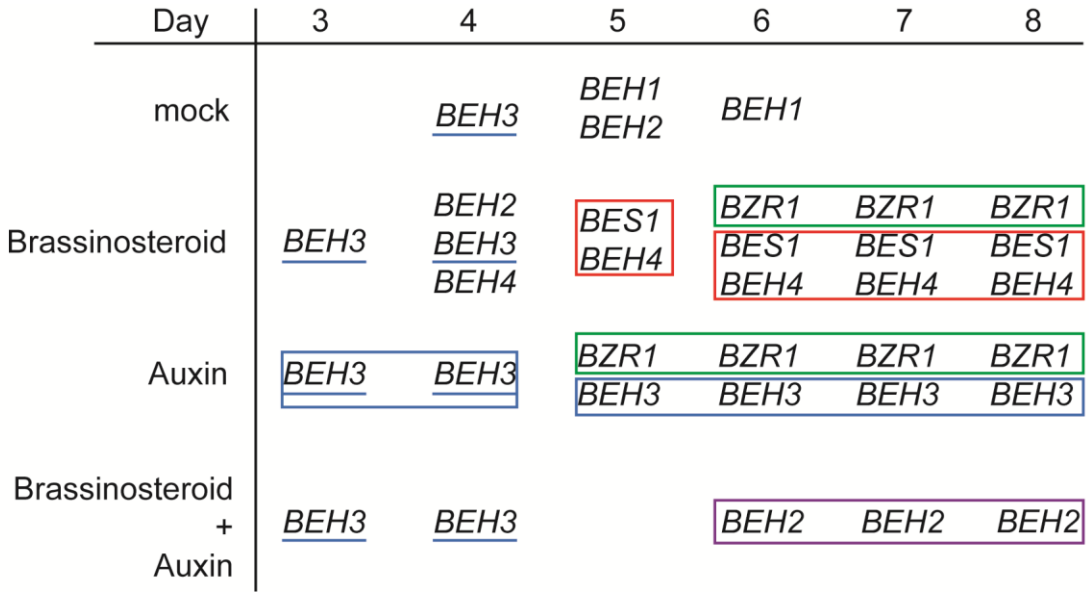

Figure 6. Degeneracy of function among *BEH* family members in hypocotyl growth. The time and condition in which a gene is needed for wild-type hypocotyl growth is listed. *BES1*, *BZR1*, and *BEH4* results are from Galstyan (pers. comm.). Degeneracy in function is found for all the time points and conditions in which more than one gene is listed. *BEH3* tends to be needed for early growth (blue underline) and response to auxin alone (blue box). *BZR1*is needed to response to single hormone treatment (green box). *BES1* is only present in combination with *BEH4* and only in response to brassinosteroids (red box). *BEH2* was the only gene exhibiting an inhibitory effect on growth in the presence of a hormone combination (purple box). All results for *BES1*, *BZR1*, and *BEH4* are from Galstyan (pers. comm.)

Supplemental Figures

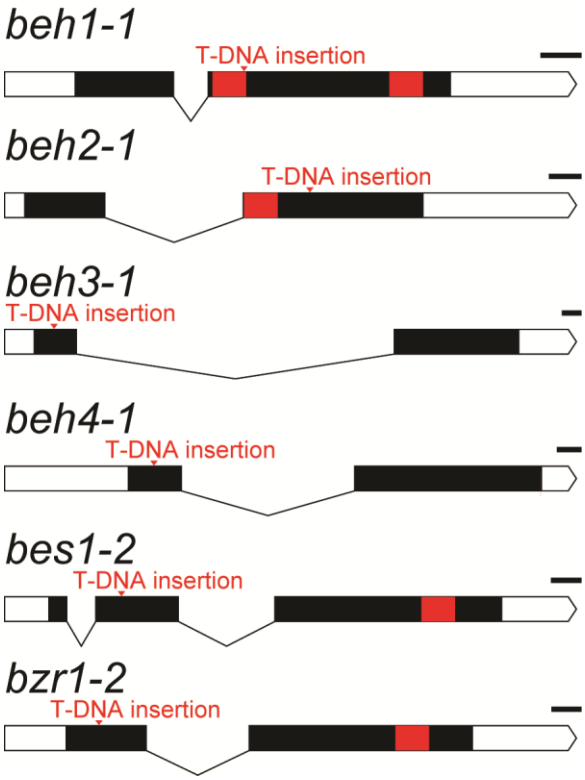

Figure S1. Location of T-DNA insertions and location of PEST domains across the *BEH* family members. PEST domains are marked in red.

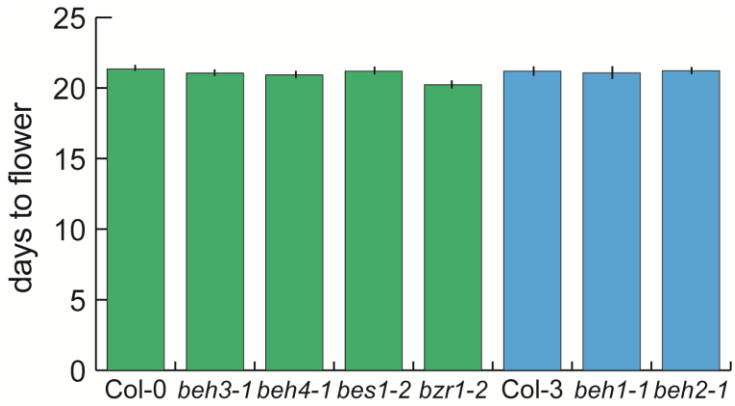

Figure S2. Flowering time is not affected in single mutants in the BEH family. Fifteen plants of each genotype were grown in long-day conditions at 22°. Flowering time was measured by the day of first emergence of the apical bud post-seeding.

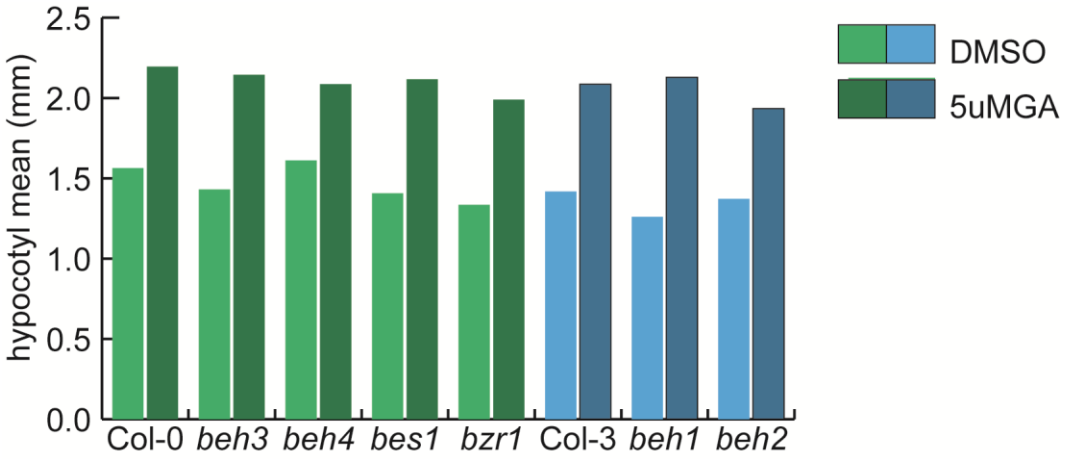

Figure S3. No difference in response to GA treatment is detected in mutants in members of the *BEH* family. Seedlings were grown on 5μM gibberelic acid (GA3) for seven days in long day conditions. Comparisons of responses between each mutant and wild-type were performed using linear regression model with genotype, GA treatment, and interaction as fixed effects.

Target gene	Oligo
BEH1 F	TTTGTCTTGAAGCTGGTTGGATCG
BEH1 R	TTCTGTTGGTCGAGAACCCTTTC
BEH2 F	TATCCAACAGTGCGCCTGTGAC
BEH2 R	AGTTTCCGCTTCGAACCACGAG
BEH3 F	TGCAATGAAGCTGGTTGGACTG
BEH3 R	TCCATTGGTTTGCATCCCTTGC
BEH4 F	GCACTCTGTAACGAAGCTG
BEH4 R	TGGCTGATAGGAAGAGCA
BES1 F	GGCTGGTTTAACTCAAATCAACGG
BES1 R	TCCGTCAGACGTCATCTTCTTCG
BZR1 F	TTGTGTTGAAGCTGGTTGGGTTG
BZR1 R	GTAAAGGCTTGCATCCCTTGCG

Table S1. Primers used for qPCR of *BEH* family members.

CHAPTER VI

MAPPING NOVEL REGULATORS OF ROOT ROBUSTNESS[1]

Introduction

For survival, all populations must strike a balance between exhibiting enough phenotypic variation to adapt to environmental changes and preserving previously adaptive phenotypes. Phenotypic robustness refers to the ability of organisms to maintain an invariant or wild-type phenotype upon perturbation (Lempe et al. 2012; Queitsch et al. 2012). A high degree of robustness is adaptive in stable environments, with the rate of environmental fluctuations dictating the ideal degree of phenotypic robustness (Pujato et al. 2013; Fraser & Kærn 2009). Phenotypic robustness has been studied in the context of three classes of perturbations: developmental noise, genetic mutations, and environmental stresses. All three of these perturbations contribute to phenotypic distributions (Hartl & Clark 2007); therefore, regulating their influence may be of utmost importance (Queitsch et al. 2012). It is thought that phenotypic robustness relies on intrinsic properties of biological networks, such as degree distribution (Whitacre 2012). It is not yet clear whether the same network features commonly provide robustness to all three of these perturbations, but there is evidence for overlap (Bergman & Siegal 2003; Lehner 2010; Queitsch et al. 2002).

[1] This chapter is in preparation for submission with co-authors Xia Shen, Christine Queitsch, and Örjan Carlborg.

The protein chaperone Hsp90 can provide robustness to all three perturbations: genetic variation, environmental stress, and developmental noise (Rutherford & Lindquist 1998; Queitsch et al. 2002; Sangster et al. 2007; Yeyati et al. 2007; Jarosz & Lindquist 2010). Hsp90's role in buffering developmental noise was examined after it was first identified as a regulator of phenotypic robustness to genetic perturbations (Rutherford & Lindquist 1998). Reduction of Hsp90's function through siRNA knock-down or pharmacological inhibition in isogenic individuals demonstrated that Hsp90 also regulates robustness to developmental noise (Queitsch et al. 2002), but this depends on the trait and organism examined (Milton et al. 2003, 2006).

Genes that regulate phenotypic robustness to developmental noise have identified using mutant screens and mapping strategies (Lempe et al. 2012). Mutant screens have identified developmental noise regulators using both morphology (Levy & Siegal 2008) and expression level (Rinott et al. 2011) as read-outs. Phenotypic robustness to developmental noise in both morphology (Hall et al. 2007; Todd A Sangster et al. 2008) and expression level (Ansel et al. 2008; Fu et al. 2009; Jimenez-Gomez et al. 2011)(Ansel et al 2008, Fu et al 2009, Jimenez-Gomez et al 2011) has been mapped in experimental populations. Recent genome-wide association studies (GWAS) have identified loci underlying variance heterogeneity. Variance heterogeneity refers to the difference in trait dispersion between two alleles across a population (Shen et al. 2012; Yang et al. 2012) and should not be confused with phenotypic robustness to developmental noise which examines trait dispersion across isogenic individuals. Although GWAS has successfully identified loci underlying variance heterogeneity of morphological characteristics, it has never been used to identify loci regulating robustness to developmental noise in a morphological trait

GWAS that have identified genes underlying robustness to developmental noise have focused on gene expression levels as phenotypes due to the ease with which a large number of individuals and genetic backgrounds can be analyzed (Ansel et al. 2008; Jimenez-Gomez et al. 2011; Nelson et al. 2013). To analyze morphological traits on the same scale, the trait in question needs to be amenable to high-throughput quantitation. *A. thaliana* seedling roots are conspicuous, easy to grow and, for all practical purposes, one-dimensional, making seedling root length an ideal trait to analyze for robustness to developmental noise for GWAS.

Seedling root development is well-studied: a small number of meristematic cells at the root tip create all of the 15 cell types of the root (Scheres et al. 2002; Cederholm et al. 2012), which include the vasculature, cortex/endodermis, columella, and epidermis. This stereotypic process arranges cells into the correct position, after which they rapidly elongate. Therefore, root length robustness arises from the spatial patterning of cell types and degree of elongation. (Scheres et al. 2002; De Lucas & Brady 2013). Several transcription factor networks have been identified that control the differentiation of various root tissue types, including the cortex and endodermis (Helariutta et al. 2000), root hair cells and non-hair cells (Pesch & Hülskamp 2004) and the stele (Brady et al. 2011). The extent of cell elongation is defined by the size and shape of the root cell walls (Scheres et al. 2002). As such, enzymes involved in the biosynthesis of plant cell walls are also required for proper root growth. Additionally, several phytohormones regulate aspects of root development. Auxin is essential for embryonic root formation, with some mutants in genes downstream of auxin lacking roots altogether (Scheres et al. 2002). Post-embryonically, auxin acts in concert with brassinosteroids to maintain root meristem identity (Mouchel et al. 2006) and with ethylene to regulate the extent of cell elongation (Stepanova et al. 2007; Swarup

et al. 2007). Variation in any of these processes can affect root length, but whether phytohormones, transcription factors, and cell wall-modifying enzymes, also regulate root robustness is unknown.

Using genome-wide association in natural accession of *A. thaliana*, we identified a locus underlying phenotypic robustness to developmental noise in seedling roots. We identified a candidate gene that may be involved in modifying phytohormones or cell wall components. Furthermore, we mapped loci underlying root mean to determine whether the same genes that underlying root robustness also regulate mean. We found that root mean was regulated by several pairs of interacting genes, including a transcription factor and a potential cell wall deposition gene. The identification of all of these loci required advanced statistical methods, as commonly used GWAS methods did not identify any of these loci. Therefore, we provide an example of a successful protocol for identifying association loci when typical GWAS fails.

Results

Heritability of root phenotypes

To map genes affecting phenotypic robustness to developmental noise, an accurate measure of trait variance is needed. Few analyses of phenotypic robustness have been completed because publically available data tend to report estimates of population means, not estimates of population variances (Atwell et al. 2010; Jimenez-Gomez et al. 2011; Lempe et al. 2012). For accurate estimates of robustness to developmental noise, large sample sizes are required to estimate variance; however, fewer than 20 individuals are typically reported for population estimates (Atwell et al. 2010). Finally, the experimental design must be able to separate the

contributions of the environment and genetic variation to determine robustness to developmental noise (Lempe et al. 2012; Geiler-Samerotte et al. 2013).

In order to estimate robustness to developmental noise, we examined the variances and means for *A. thaliana* root length across 93 accessions with a large degree of replication (Figure S1, Table S1). The variance for each accession was independently calculated from three biological replicates with n = 70 (Figure S1). As *A. thaliana* is an inbreeding species uncontrolled genetic variation is virtually eliminated. We grew seedlings on a standard plant medium in temperature-controlled chambers to reduce environmental variation among individuals. Such an experimental design allows for parsing of the genetic and environmental factors from the developmental noise underlying root phenotypes. We examined the reproducibility of variance and found no significant differences among the three replicates (Figure 1a, ANOVA, $p = 0.079$).

Mapping the loci underlying variation in robustness to developmental noise requires a difference in that trait among the accessions. We calculated the means from the repeated measures of variance and determined that the variance varied significantly across the accessions ($p = 7.5*10^{-239}$, Brown-Forsythe test), suggesting a genetic basis for the variance differences that could be identified by genetic association mapping. We also estimated the pooled variance by combining the measurements from the three biological replicates after correcting for replicate effects (see Methods) to estimate variance with even greater accuracy. We found a strong correlation between the mean variance of the three biological replicates and the pooled variance ($p = 1.8*10^{-68}$, $R^2 = 0.97$, Figure S2a), so we used the pooled variance measurements in further analyses.

Previous studies that identified loci associated with phenotypic robustness to developmental noise have found an overlap with the loci associated to the trait mean (Hill et al. 2008; Jimenez-Gomez et al. 2011). This potential confounding effect of the mean is a concern in examining phenotypic robustness (Geiler-Samerotte et al. 2013) as the classical statistical models operate on the assumption that the variance of traits is independent of the mean. If this assumption is violated, other models must be used. We therefore examined root length means (Figure 1b) and found that the accessions varied significantly in root length ($p < 2.2e\text{-}16$, ANOVA). This also showed that root length mean has a genetic basis that could potentially be mapped. As there was no correlation between root length mean and root length variance in our data ($p = 0.11$, $R^2 = 0.02$, Figure S2b), the phenotypic robustness to developmental noise across the accessions was not confounded by the root mean in our dataset. Hence we could explore the potential overlap in the genetic architectures underlying root length variance and root length mean without concerns about confounding between the two.

To map loci underlying a trait, the trait needs to be under genetic control. It is possible that variance, unlike mean, is a stochastic trait with no underlying genetic cause. We calculated the heritability of the pooled means and variances, while correcting for the population structure among the accessions, using a linear mixed model implemented in the R/hglm package (Rönnegård et al. 2010) (Table 1), which simultaneously considers the mean and the dispersion of the data. The broad sense heritability was 0.33 for root mean and 0.02 for root variance. The linear mixed model also partitions the variance explained into the additive effects and the

epistatic contribution. For the root mean, a larger fraction of the variation was explained through epistatic effects (0.21) than by additive effects (0.12).

Root length variance is associated with SNV in the gene At3g46700

Despite the low broad-sense heritability, we performed a genome-wide association analysis for phenotypic robustness to developmental noise across the *A. thaliana* genome. The 93 phenotyped accessions were all previously genotyped at ~215,000 loci (Atwell et al. 2010). We used an additive model for GWAS and corrected for population structure by the principle components of the genomic kinship matrix ("egscore" procedure in R/GenABEL), based on the ~215,000 genotyped SNPs across the accessions. No SNPs were found to be significantly associated with root variance (Figure S3), but one SNP almost reached the conservative Bonferroni-corrected significance threshold used here. To analyze this potential association more thoroughly, we used a whole-genome generalized ridge regression heteroscedastic effects model (HEM), in which SNPs are included as random effects and their effects were estimated simultaneously using the R/bigRR package (Shen et al. 2013). A 5% genome-wide significance threshold was determined via 1000 permutations of the phenotypic vector. In this analysis, a significant association was mapped (Figure 2a) to a SNP on chromosome 3 at position 17,201,307bp (3_17201307) where the minor allele was associated with greater root variance (Figure 2b).

The Linkage Disequilibrium (LD) decays within 10kb in *A. thaliana* (Kim et al. 2007). The SNP associated with root mean (3_17201307) was not in LD ($R^2 > 0.8$) with any other SNPs within 30kb, suggesting that the causal variant is located very close to this top SNP.

Interestingly, 3_17201307 encodes a missense variant in the gene At3g46700, which is a member of the UDP-glycosyltransferase superfamily based on sequence homology (Lamesch et al. 2012). UDP-glycosyltransferases catalyze the enzymatic addition of sugars onto a variety of other molecules, such as secreted components of the plant cell wall (Scheible & Pauly 2004) and phytohormones (Husar et al. 2011; Poppenberger et al. 2005; Priest et al. 2005; Jin et al. 2013; Kleczkowski et al. 1995). However, the substrates of At3g46700 have not been identified. The associated variant (gCg→gAg) results in a polar glutamate amino acid replacing the non-polar amino acid alanine, which has the potential to alter protein structure. The missense change is not found in the UDP-glycosyltransferase domain, suggesting that the mutation may change the target specificity or efficiency of the protein. At3g46700 is most highly expressed in adult roots in seedlings, making it a strong candidate functional mutation for the observed association (Schmid et al. 2005). Two different stop codons are present in At3g46700 in several accessions (Cao et al. 2011), indicating that loss of function mutations are not lethal. These two stop codon are not in LD with the leading SNP. We compared the root variance among accessions containing a stop codon versus accessions without stop codons. The variance tended to be higher among accessions with a functional At3g46700 ($p=0.06$, t-test), further suggesting that altered function of At3g46700 might be involved in regulating root variance.

At3g46700 directly controls root phenotypic robustness

In order to partially alter the activity of At3g46700, we reduced At3g46700 expression using an artificial microRNA (amiRNA) (Carbonell et al. 2014). Sequences that are predicted to target At3g46700 were inserted into an amiRNA vector, and three different amiRNAs targeting At3g46700 were created (Table S2). To verify that the amiRNA was targeting At3g46700,

expression analysis of At3g46700 and expression of the amiRNA will be completed. 5' RACE will also be used to verify cleavage of At3g46700. With validated lines, variance will be measured in three independent lines for each of the amiRNAs and the vector control.

Genetic architecture underlying root mean

To explore the genetic architecture of root mean, we performed a GWAS as described above with root mean instead of variance as the phenotype. No significant signals were found after multiple testing correction, neither in the standard GWAS (Figure S4a) nor the HEM analyses (Figure S4b). This is rather surprising, given that the narrow-sense heritability of this trait is 0.12. However, as much of the variability underlying root mean is epistatic (Table 1), we decided to also explore the loci with effects beyond additivity.

It is generally thought GWAS is underpowered to detect epistasis (Gibson 2012). Epistatic models are commonly examined in experimental populations, such as RILs (Mackay 2001), but examinations of epistasis in natural population are rare (Carlborg et al. 2006; Hemani et al. 2014). GWAS studies considering epistasis generally test for interactions between candidate genes identified using an additive model or genes known to act in the same biological processes (Pyun et al. 2013; Sapkota et al. 2013; Hemani et al. 2014). Using all variants to scan for epistatic effects has many challenges including difficulty in finding low frequency two-locus genotype classes and the need to use stringent significance-thresholds to account for the large number of statistical tests performed. Together these challenges mean that epistatic GWAS analyses generally result in few significant combinations. In spite of these difficulties, we felt that an unbiased SNP-by-SNP epistasis analysis (PLINKv1.07, (Purcell et al. 2007)) was

motivated since our initial analyses indicated that non-additive genetics makes a significant contribution to the root mean phenotype in this population (Table 1) and that the relatively small genome of *A. thaliana* decrease the number of pairwise combinations to test. To maximize power, we only tested for epistasis among SNPs with a minor allele frequency greater than 25% to increase the chances that all allele combinations would be present for the tested SNP by SNP combinations. All SNP by SNP results with fewer than four accessions in the minor genotype class were removed and population stratification was accounted for using a linear mixed model with genomic kinship matrix as in the single-locus GWAS analyses. The significance threshold was calculated by estimating the number of independent tests from the number of linkage blocks across the *A. thaliana* genome (Kim et al. 2007). The *A. thaliana* genome is approximately 125Mb with average LD-blocks of about 10kb. In total, the epistatic analysis thus included approximately 78 million independent tests and using a Bonferroni-threshold accounting for this number of tests, a total of four interacting pairs were significant (Figure 3a, Table 2, Figure S5). These four interactions explained 0.19 of the variance in root length, which is nearly the total estimated epistatic contribution (0.21) (Table 1). We examined the root lengths and the frequencies for each genotype class for the four significant pairs (Figure 3b). The same pattern was observed for each SNP by SNP combination: the genotype class with the longest mean root-length was present in the lowest frequency. Although this type of epistasis contributes with a marginal additive effect, the low frequency of the high-mean two-locus genotype means that the contribution to the additive variance was very low. This is a likely explanation of why these loci were not detected in the standard GWAS analysis.

As the genotype→phenotype map for the associated SNP-pairs were similar, we tested whether this could be due to LD between them. Except for the pairs 3_9272294, 3_9273674 and 5_15862026, 5_15862525, which are within 1.5kb and 505bp respectively, the SNPs were independent ($R^2 < 0.8$) from each other. In an attempt to fine-map the significant associations, we also explored the 13 million SNPs obtained by whole-genome re-sequencing in the 1001 Genomes effort (Ossowski et al. 2008). SNPs that were in linkage disequilibrium ($R^2 > 0.8$) and located within 5kb upstream or downstream of the leading SNP in were considered further (Table 2). For each pair, the effects of the SNPs that fulfill these criteria together with their locations are listed in Table 3. Below, we discuss the four detected pairs in more detail.

Pairs 1, 4

A SNP on chromosome 3, 3_66596, was part of two of the four significant interacting SNP pairs, suggesting a potentially important role of this locus in the regulation of root length. It was not in LD with any other SNPs, suggesting that it is located on a short LD-block. Interestingly, the SNP itself results in a synonymous change in the gene At3g01185, a 148 amino acid protein with unknown function (Lamesch et al. 2012). It is also located within 500bp of the transcriptional start site of At3g01180, where it co-localizes with a peak in DNase accessibility (plantregulome.org, March 2014), suggesting that it may also affect the expression of this downstream gene. At3g01180 encodes *STARCH SYNTHASE 2* (*SS2*), which is involved in the storage of glucose as amylose. *ss2* single mutants have no effect on aerial tissue development or the amount of starch stored, but instead alter starch composition: the amylopectin chains become shorter (Santelia & Zeeman 2011). Across plant development *SS2* was not highly expressed in the seedling root, though within the root phloem, *SS2* expression appears to oscillate during its

development (Winter et al. 2007). To explore whether 3_66596 affects root mean via effects on At3g01180 or At3g01185, we measured the root length of T-DNA insertion lines with those genes interrupted. The T-DNA insertion line affecting At3g01185 exhibited no difference in root growth from wild-type (Figure 4), implying that alteration of this gene in isolation is not central for the regulation of root length. The T-DNA line affecting At3g01180 will be tested once creation of the line is completed.

Two SNPs in high LD within 2kb of one another, 3_9272294 and 3_9273674, interacted with SNP 3_66596. In total, four genes were affected by the 21 SNPs that were in high LD with the two leading SNPs (Table 3). The first gene, At3g25545, has three SNPs in its putative promoter. At3g25545 is an unstudied protein; however, it is found to be expressed in the vasculature of seedling roots and imbibed seeds (Winter et al. 2007). The second gene, At3g25540, was affected by 5 different SNPs, including a missense variant. At3g25540 encodes a ceramide synthase *LAG ONE HOMOLOG* (*LOH1*), which in combination with LOH3 is essential for sphingolipid synthesis (Markham et al. 2011). Sphingolipids can be involved in cell wall deposition (Chen et al. 2008). In the *loh1-1* single mutant, no developmental defects were noted in Markham et al (2011); however, in *loh1-1 loh3-1* double mutants, root growth was abrogated on soil. Notably *LOH1* is most highly expressed in the seedling root and imbibed seed relative to other tissues (Winter et al. 2007). The missense variant is not located in the conserved domain of *LOH1*, the TLC transmembrane domain (Lamesch et al. 2012; Finn et al. 2006); rather the variant is located in an unexamined region of the protein. These findings suggest this missense variant could affect the function of LOH1, thus altering root length. The third candidate gene was At3g25530, which was affected by a large number of variants (Table 3). At3g22530

encodes *GR1*, a glyoxylate reductase, known to be involved in maintaining redox balance in response to stresses (Breitkreuz et al. 2003; Allan et al. 2008). This gene is not highly expressed in the root; rather it is primarily expressed in leaves and cotyledons (Winter et al. 2007). The missense variant neither affected the NAD binding domain of *GR1*, nor was the S/T change predicted to be phosphorylated (PhosPhat4.0, (Zulawski et al. 2013)). The final candidate gene At3g25520 encodes ATL5/OLI5/PGY3, a 5S rRNA binding protein. Three SNPs were found in the promoter of *ATL5* and one found in an intron, suggesting that these SNPs could alter the regulation of this gene. *ATL5* is expressed in seedling roots (Winter et al. 2007), and the loss-of-function mutant *oli5-1* exhibits shorter primary roots than wild-type (Fujikura et al. 2009).

Based on previous studies and publicly available expression data, several genes in the associated region are thus functional candidates for the root length association. The strongest candidate appears to be *ATL5*, but more work is needed to explore which one of these is the interactor with the genes underlying the association to SNP 3_66596.

To determine which of the functional candidate genes regulate root length, we examined the mean root length in T-DNA insertion mutants for At3g25530 and At3g25540. No significant difference in root length was observed for At3g25530 or At3g25540 insertions relative to wild-type (Figure 4). I will test the T-DNA insertion in At3g25520. I also will explore the possibility that variants in LOH1 could also regulate root mean length by crossing the T-DNA insertion in At3g25530 with the At3g01180 and At3g01185 T-DNA insertion lines.

One other SNP on chromosome 5, 5_18241640, also interacted with SNP 3_66596. This SNP was in LD with two missense SNPs and four synonymous SNPs, all located in the gene At5g45120. At5g45120 encodes a member of the little studied eukaryotic aspartyl protease

family (Beers et al. 2004). At5g45120 is expressed primarily in pollen and maturing seeds. The two missense variants, I→T and D→A are located in the aspartyl protease domain. The missense variants are positioned 5bp apart and are either adjacent to or within a 'phosphorylation hotspot', suggesting that post-translational modifications may be altered in certain accessions (Heazlewood et al. 2008; Durek et al. 2010). We found that interrupting At5g45120 with a T-DNA insertion did not affect root length (Figure 4), suggesting that the effect of At5g45120 might be dependent on a specific variant at 3_66596 to express its effect. To explore this I will cross the line with the T-DNA insertion in At5g45120 with the lines with a T-DNA insertion in At3g01180 and At3g01185.

Pair 2

A SNP on chromosome 1, 1_17257526, interacted with two SNPs in close LD on chromosome 5, 5_15862026 and 5_15862525. The SNP 1_17257526 was not in LD with any others SNPs and was located in the gypsy-like transposon At1g46624. The 20kb region surrounding 1_17257526 encodes 16 transposable elements, and has accessible chromatin (plantregulome.org, March 2014). How a polymorphism in this region would affect root length by itself, or through interactions, is unclear. The SNPs 5_15862026 and 5_15862525, and other SNPs in high LD with them, were all intergenic with the nearest gene 1.9kb away. This region does co-localize with accessible chromatin (plantregulome.org, March 2014), suggesting that it may be regulatory. The nearest downstream gene was At5g39610, the transcription factor *NAC6*. *NAC6* regulates leaf senescence (Aeong Oh et al. 1997) and is highly expressed in senescing leaves and in maturing seeds (Winter et al. 2007). We examined a T-DNA insertion line in which At5g39610 is interrupted. We found a significant reduction in root length in the knock-out

mutant (Figure 4), indicating that altered regulation of At5g39610 could affect root length. Further work is needed to explore how this gene could potentially interact with the transposon on chromosome 1.

Pair 3

The third interacting pair includes SNPs on chromosome 3 and 5 (3_10891195 and 5_1027939). SNPs in LD with 3_10891195 are located in three genes (At3g28865, At3g28870, and At3g28880). At3g28865 is a LINE retrotransposon that has not earlier been associated with root length. At3g28870 is affected by a missense variant and a synonymous variant. This gene is unstudied, but contains a number of conserved domains with the missense variant located within the Sec63 domain. In yeast, Sec63 is needed for the assembly of ER translocons (Jermy et al. 2006). At3g28870 is found most highly expressed in maturing seeds and adult leaves (Winter et al. 2007), and it also is expressed in pollen, along with a number of other unstudied proteins that share conservation (Holmes-Davis et al. 2005). The last candidate in the region, At3g28880, encodes an unstudied ankyrin family protein for which there is no information on developmental expression patterns. The SNPs affect the promoter of At3g28880. We examined a T-DNA insertion mutant for At3g28880 and found that there is a significant reduction in root length. In this region, at least one gene At3g28880 affects root length mean. I will also test At28870 for its root length once the T-DNA line is obtained (Figure 4).

The SNP on chromosome 5, 5_1027939, is intergenic and located 579bp upstream of At5g03840. At5g03840, *TERMINAL FLOWER 1* (*TFL1*), is highly expressed in both the adult and seedling root (Winter et al. 2007). *TFL1*'s role in floral initiation and morphology have been well studied in many plant species (Jack 2004) and it also has a role in regulating the storage of

proteins in vegetative tissues, such as roots or seeds (Sohn et al. 2007). *tfl1* roots, however, develop with wild-type morphology (Larsson et al. 1998). Another potentially affected gene upstream of 5_1027939 and its single linked SNP is At5g03850. At5g03850 encodes a nucleic acid-binding, OB-fold-like protein, which is expressed in seedling roots. I will also examine lines with a T-DNA insertion in At5g03840 and At5g03850 for root length (Figure 4).

In total, 13 genes are found in linkage to regions associated with root length through epistatic interactions. In addition, two transposons are also found in these regions. So far, seven of these genes have been tested for root phenotypes when interrupted by a T-DNA insertion. In two cases, root length was affected. As of writing, five additional T-DNA lines are ready for analysis, with two having support in the literature for affecting root length. With the continuing validation assays, we will be able to judge whether mapping interactions underlying root length was successful.

Discussion

GWAS have identified loci affecting the phenotypic robustness to developmental noise using gene expression as a read-out (Ansel et al. 2008; Jimenez-Gomez et al. 2011; Nelson et al. 2013); however, in this study we performed the first GWAS for robustness to developmental noise of a morphological trait. We assessed a large number of individuals to estimate variance accurately compared to studies performing GWAS on mean traits, which require less replication for accurate estimates. In spite of the low heritability of this trait, we were able to map an associated locus by using advanced statistical methods to detect association

We identified a variant in At3g46700 as associated with developmental noise in root length. At3g46700 encodes a UDP-glycosyltransferase, a member of a family of enzymes that add sugar groups to a variety of substrates (Li et al. 2001). Several potential substrates of At3g46700 are known to be important to root length, such as cell wall components and phytohormones (Scheible & Pauly 2004; Husar et al. 2011; Poppenberger et al. 2005; Priest et al. 2005; Jin et al. 2013; Kleczkowski et al. 1995). The 25kb region flanking At3g46700 also encodes 6 duplicated copies of a UDP-glycosyltransferase with the same exon-intron structure as At3g46700. At3g46700 shares between 59-73% amino acid identity with these other UDP-glycosyltransferases (Johnson et al. 2008). Among the most identical UDP-glycosyltransferases adjacent to At3g46700, two were also highly expressed in seedling roots (Winter et al. 2007). Redundancy and degeneracy can provide robustness to traits (Wagner 2005). Therefore, changes to the shared function of these similar UDP-glycosyltransferases due to the associated variant in At3g46700 could contribute the change in robustness observed.

The genetic architecture underlying root length mean differed greatly from that of root length variance. First, the genetic contribution to root length variance was very low. Second, additive genetic variation explained some of root length mean, but a greater contribution of interacting genes was detected. We were unable to map loci that affecting root length mean additively. Instead a gene-by-gene GWAS study implicated four genetic interactions in affecting root length mean. The effect size of these four interactions explained nearly all (0.9) of the epistatic contribution to the genetic variation.

One of the significantly interacting loci associated with root mean, 3_66596, was present in two of the significant interactions, suggesting that it is an important regulator of root length.

One of the loci interacting with 3_66596 encodes two candidate genes that have known roles in regulating root length, *ATL5* and *LOH1*. The significant signal we detect in the gene by gene GWAS may be due to both genes in the region being causal. The last of the interactions that include 3_66596 includes a gene encoding an aspartyl protease family protein. The T-DNA insertion line showed no effect on root length demonstrating that the protease alone cannot regulate root length but may do so in concert with variation at 3_66596.

For one of the pairs, we identified variation in the promoters of a transcription factor that regulates root length in concert with variation at the interacting locus. A T-DNA insertion in the transcription factor confirmed that *NAC6* affects root length. Previously, *NAC6* had been identified as a positive regulator of plant senescence, increasing in expression among aging rosette leaves (Im Kim et al. 2011). Here, we have identified a new role for this transcription factor in root growth regulation. The locus that contains genetic variation interacting with *NAC6* variation encodes a transposon. Transposons have been previously found to influence phenotypes *A. thaliana*. For example a transposon is present in the intron of *FLC* in some accessions, altering *FLC* expression (Gazzani et al. 2003). However the transposon we have identified is located in an intergenic region on a different chromosome than *NAC6*. How this transposon could affect phenotypes in a way dependent on *NAC6* variation is unclear.

Several of the candidate interacting SNPs are found in non-coding regions. For example, we identified an interaction between a SNP in the promoter of the gene encoding an unstudied ankyrin-repeat containing protein this gene and SNPs located in a non-coding region of the genome. Therefore, the effect on root length mean due to these interacting loci is likely at the level of gene expression. This is not the first example of an interaction between non-coding

regions. For instance, interaction between variation in the promoters of FCRL3 and NFκB1 is associated with the occurrence of rheumatoid arthritis (Martinez et al. 2006).

This study demonstrates the utility of non-standard GWAS analyses for identifying loci underlying traits of interest. A candidate locus that was trending towards significance in standard GWAS surpassed multiple-testing threshold once methods that use alternative approaches for dealing with multiple tests and trait variation. Furthermore, knowledge of the types of genetic contribution informed our methods. The genetic component of root length mean was largely epistatic, guiding us to complete an unbiased GWAS of interactions. This method proved extremely successful, enabling the identification of virtually all the epistatic contribution to the trait. Future GWAS should consider utilizing such methods to improve the identification of causal genes.

Methods

Phenotyping

In order to reduce environmental variation among accessions, eighteen individuals from each of the 93 accessions studied in Atwell et al (2010) (Table S1) were vernalized at 4° as seedlings for six weeks to synchronize growth and flowering. The five most developmentally advanced seedlings from each accession were then transferred to soil in a randomized design. Plants were grown in 1ong-day conditions at 22°. Flats were rotated three times per week to reduce position effects on plant development. Seeds were collected over a period of three months as the plants dried. Equal numbers of seeds were pooled from 3-5 parent plants to reduce the environmental contribution of plant placement across flats of plants. Ethanol sterilized seeds

were planted on 1× Murashige and Skoog (MS) basal salt medium supplemented with 1× MS vitamins, 0.05% MES (wt/vol), and 0.3% (wt/vol) phytagel in a semi-randomized design with n = 70 per accession. Four sets of twenty-three accessions plus a control accession (Col-0) made up a set. Each set was replicated three times providing the standard error for variance and CV, with a total n = 210 planted for each accession. The seeds were stratified at 4° for three days and then grown for seven days in darkness with the plates in a vertical position. A photograph of each plate was taken, and the hypocotyl and root means were measured manually using the 'freehand' function in ImageJ1.46r (Schneider et al. 2012). Non-germination, missing organs, and delayed development were noted.

Data 'standardization'

Thirty-one seedlings with hypocotyls less than 5mm were removed because it is likely that germination was severely delayed (Todd A Sangster et al. 2008). Between-set differences were corrected by subtracting the difference between Col-0 in each replicate and set and the global mean for Col-0. Systematic differences between replicates were still present and corrected within each set by using the residuals from a model in which hypocotyl length or root length were explained by the within-set replicates.

Heritability predictions

Heritability of root length variance and mean were estimated using the repeated measures with genotype as a fixed effect using ANOVA. To parse the contribution of additive and epistatic effects, a linear mixed model including both additive and epistatic effects as random effects was fitted using the R/hglm package (Rönnegård et al. 2010), i.e.

$$y = \mu + a + b + e$$

where

$$a \sim N(0, G\sigma_a^2)$$

are the additive genetic effects, and G is the genomic kinship matrix;

$$b \sim N(0, G \circ G\sigma_b^2)$$

are the epistatic effects, and G is the genomic kinship matrix; e are independent and normally distributed residuals. The narrow sense heritability was estimated as the ratio of the additive genetic variance to the total variance, and the broad sense heritability was estimated as the ratio of the sum of the additive and epistatic variance to the total phenotypic variance.

GWAS analysis

Using the transformed mean and variance estimates and the genotypes for the 93 accessions from Atwell *et al* (2010), we performed a series of analyses to detect associations. To control for population structure in the analyses, we used the function egscore from R/GenABEL (Aulchenko et al. 2007). Next, we ran the function bigRR and bigRR_update in the R/bigRR package (Shen et al. 2013) in which the SNP effects are modeled simultaneously as random effects and the effects were estimated via the generalized ridge regression method HEM. The maximum absolute effect sizes in 1000 permutations were extracted to determine the 5% genome-wide significance threshold, i.e. 0.0020.

To screen the genome for pairs of epistatic SNPs, we used the PLINK –epistasis procedure (Purcell et al. 2007) that is based on the model:

$$Y \sim b_0 + b_{1.A} + b_{2.B} + b_{3.AB} + e,$$

which considers only allele by allele epistasis in $b_{3.AB}$. In the analysis, all possible pairs of SNPs with a minor allele frequency > 25% were tested.

We next filtered out the pairs where there were fewer than 4 accessions in the minor two-locus genotype-class. Further, all combinations in which the significance was $< 1*10^{-8}$ were also removed as it was considered unlikely that they would be significant after correction for population structure and applying a multiple-testing corrected significance threshold. For the remaining pairs, a linear mixed model including fixed and additive and epistatic effects, as in the PLINK based initial scan, and a kinship correction for population stratification, as in the single locus analyses, was fitted by R/hglm. The multiple-testing corrected significance threshold for this epistatic analysis ($3.2*10^{-10}$) was obtained by estimating the number of independent tests based on the number of estimated LD blocks in genome (12,500; (Kim et al. 2007)) and applying a Bonferroni correction for the 78 million tests performed. Using data on the sixty-three accessions from the GWAS that were unambiguously identified in the 1001 Genomes data (Cao et al. 2011), we then identified additional SNPs in LD with the leading SNP using the function *LD* in the *genetics* package in R across a region of 10kb around the marker.

Examination of SNP-associated genes

The Ensembl Variant Effect Predictor based on the TAIR10 release of the *A. thaliana* genome was used to determine the effects of the leading SNPs and the SNPs in high LD with

them. Genes were considered as candidates if they were within 1kb of a variant. Expression of the candidate genes was determined using the BAR eFP Arabidopsis browser (Winter et al. 2007).

amiRNA generation and transgenic line creation

amiRNA targets sequences were created using a web-based tool and ligated into the pMDC32B-AtMIR390a-B/c as described in (Carbonell et al. 2014). Agrobacterium strain GV3101 was transformed with the amiRNA plasmids. Plants were agrobacterium transformed as described in (Logemann et al. 2006).

Validation of interacting loci

T-DNA lines were obtained for the candidate genes (Table S3). Root length was ascertained as described above (n = 20). Tukey's HSD post hoc test was used to compare root lengths across the T-DNA lines and wild-type (Col-0).

Main Figures

	hglm			ANOVA
root length	total	additive	epistatic	H^2
Mean	0.33	0.12	0.21	0.25
Variance	0.02	0.01	0.01	NA

Table 1. Heritability of root length mean and root length variance. Heritability was estimated using a linear mixed model including both additive and epistatic effects as random effects using the R/hglm package. Root length mean is heritable with a large epistatic contribution, whereas root length variance has a low heritability. A similar H^2 was estimated for root length mean using ANOVA with genotype as a fixed effect.

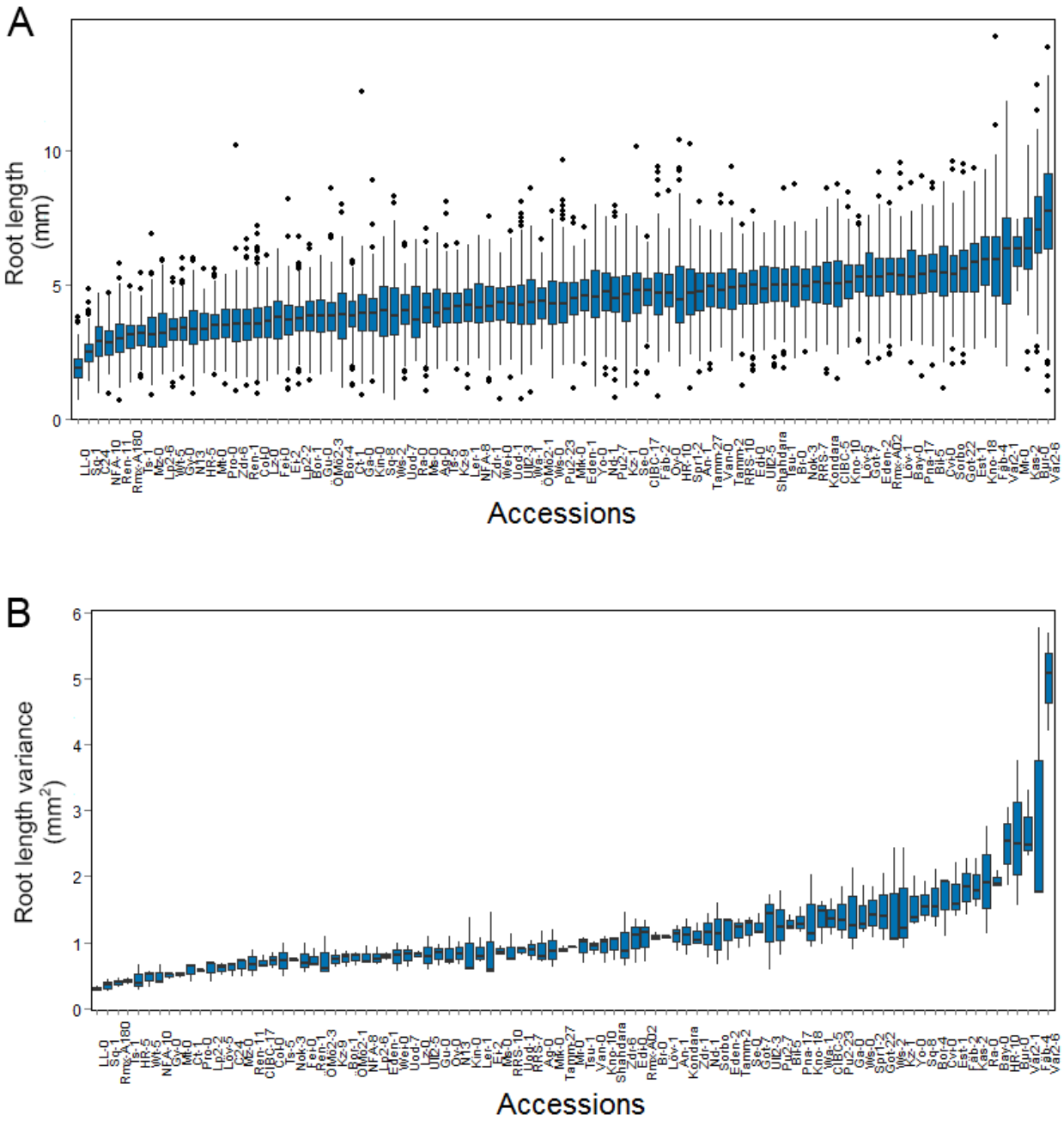

Figure 1. Across accessions, there is a broad distribution in root length and variance. A) Root length was measured in ~210 dark grown seedlings at seven days. B) Root length variance was measured in three replicates of 70 individuals.

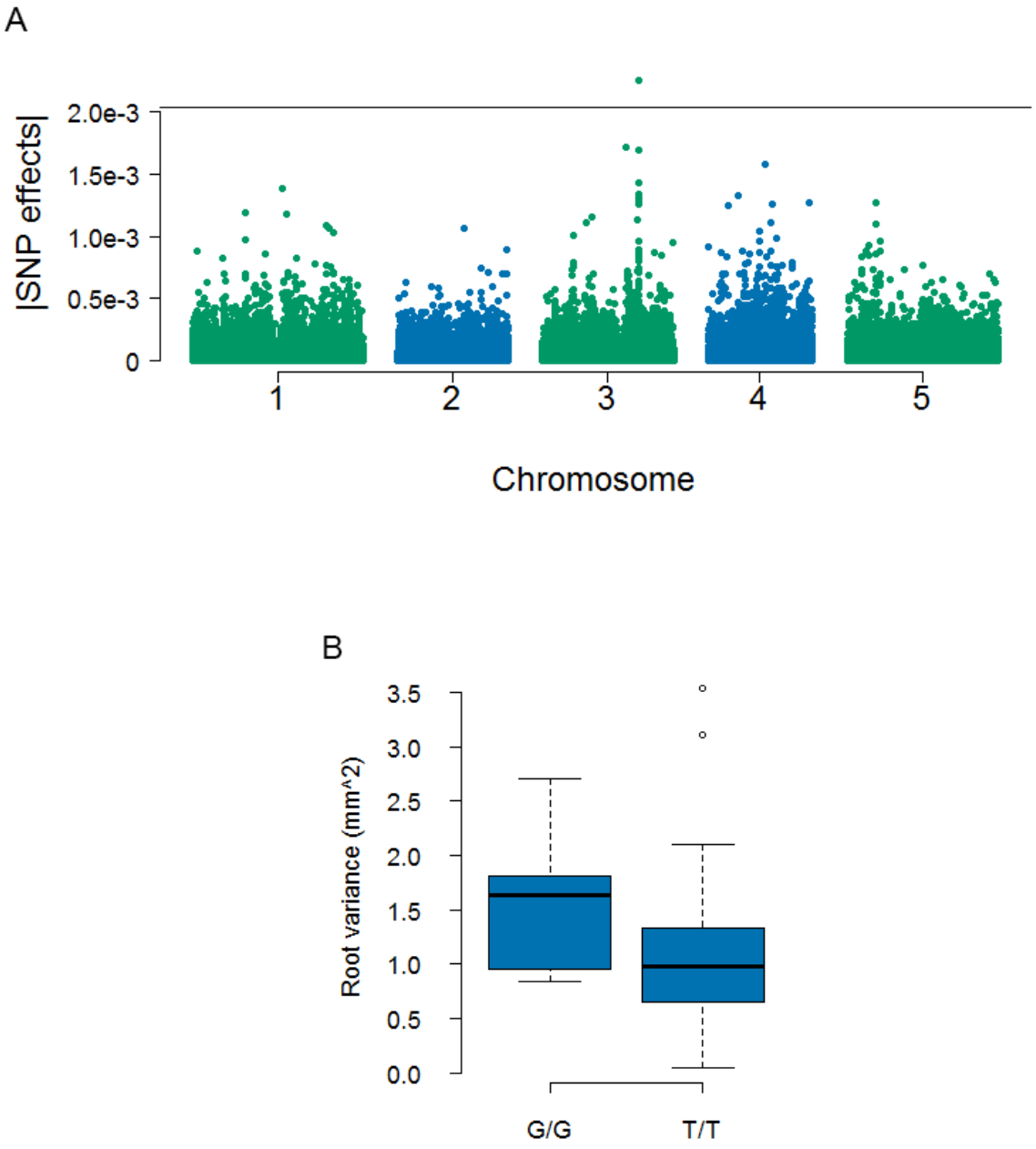

Figure 2. A variant on chromosome 3 is associated with root variance. A) Using a whole-genome generalized ridge regression model in which SNPs were random effects, a significant SNP is detected in association with root variance. B) The minor allele (G) is found associated with greater root length variance.

Pair	SNP1	SNP2	Interaction effect size (mm)	p-value
1	3_66596	3_9272294	0.5548	1.48E-10
1	3_66596	3_9273674	0.5548	1.48E-10
2	1_17257526	5_15862026	0.6337	1.67E-10
2	1_17257526	5_15862525	0.6337	1.67E-10
3	3_10891195	5_1027939	0.6256	2.25E-10
4	3_66596	5_18241640	0.6251	3.03E-10

Table 2. Six significant interactions were associated with root length mean.

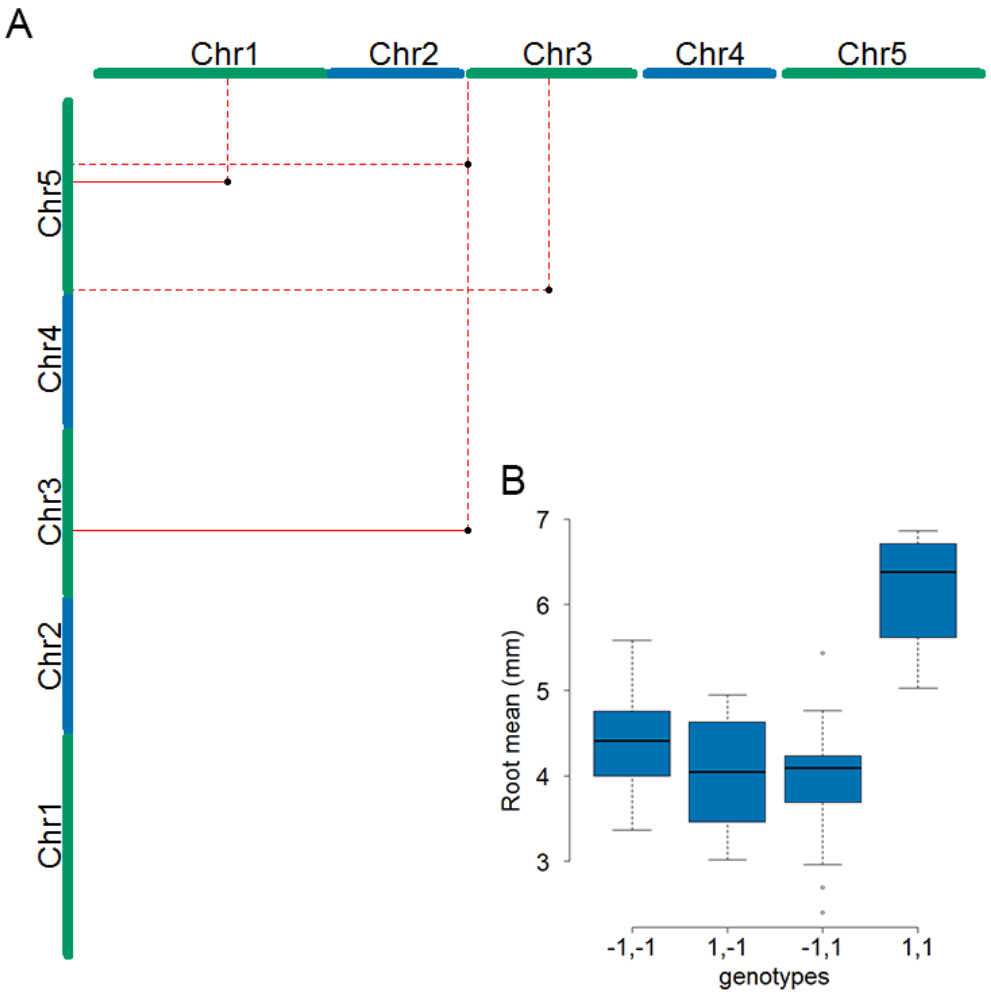

Figure 3. Four distinct genetic interactions are associated with root mean. A) The five *A. thaliana* chromosomes make up the x- and y-axes. The positions of interacting SNPs are indicated by a black point. Solid lines indicate support for an interaction by more than one linked SNP and dotted lines indicate support by a single SNP. B) Pictured is a representative example of the root mean differences among the four genotype combinations from the interaction between 1_17257526 and 5_15862026 /5_15862525. The major allele is indicated by -1, and the minor allele is indicated by 1.

Pair	Chr	Position (bp)	# linked mutations	Locus	#Mut	Mutation Effects
1,4	3	66596	0	At3g01180	1	1 upstream
				At3g01185	1	1 syn
1	3	9272294/9273674	21	At3g25520	4	1 intron, 3 upstream
				At3g25530	14	1 mis, 1 splice, 1 syn, 2 intron, 9 upstream
				At3g25540	17	1 mis, 4 syn, 2 intron, 5 5'UTR, 5 upstream
				At3g25545	3	3 upstream
4	5	18241640	5	At5g45120	6	2 mis, 4 syn
2	1	17257526	0	At1g46624	1	1 non-coding exon
2	5	15862026/15862525	4	At5g39610	1	1 upstream*
				At5g39620	1	1 downstream*
3	3	10891195	7	At3g28865	4	4 non-coding exon
				At3g28870	2	1 mis, 1 syn
				At3g28880	3	3 upstream
3	5	1027939	1	At5g03840	1	1 upstream
				At5g03850	1	1 downstream

Table 3. Variants in linkage with leading interacting SNPs from whole genome interaction analysis. Six interacting pairs of loci were found significantly associated with root length, as displayed in Table 1. The genes within 1000bp of the linked SNPs are listed and the mutation mutational effects are labeled: syn = synonymous variant, mis = missense variant, splice = splice site variant. #Mut refers to the number of mutations associated with each gene, and some mutations are associated with more than one gene. *SNPs more than 1000bp away from the nearest gene.

Pair	Chr	Position (bp)	Locus	Gene/Putative function	Highest expression tissues
1,4	3	66596	At3g01180	*SS2*	cauline leaves
			At3g01185	unknown	NA
1	3	9272294/ 9273674	At3g25520	*ATL5*	apical meristem, seedling root
			At3g25530	*LOH1*	rosette leaves
			At3g25540	*GR1*	seedling root, imbibed seed
			At3g25545	unknown	seedling root (vasculature), imbibed seed
4	5	18241640	At5g45120	Eukaryotic aspartyl protease family protein	adult root (meristem), imbibed seed
2	1	17257526	At1g46624	gypsy-like retrotransposon	NA
2	5	15862026/ 15862525	At5g39610	*NAC6*	seedling root (vasculature)
			At5g39620	RAB GTPase homolog G1	adult root vasculature, mature seed
3	3	10891195	At3g28865	LINE retrotransposon	NA
			At3g28870	Histone deacetylase interacting domain	mature seed
			At3g28880	Ankyrin repeat family protein	NA
3	5	1027939	At5g03840	*TFL1*	adult root vasculature (mature zone)
			At5g03850	Nucleic acid-binding, OB-fold-like protein	seedling root, apical meristem

Table 4. Function and expression patterns of genes in linkage with leading SNP from whole genome interaction analysis. The gene names or proposed function is listed for all genes harboring variants in linkage with the leading SNPs found to interact in the whole genome interaction analysis. Locations of high expression were obtained from the BAR eFP *Arabidopsis* Browser (Winter et al. 2007).

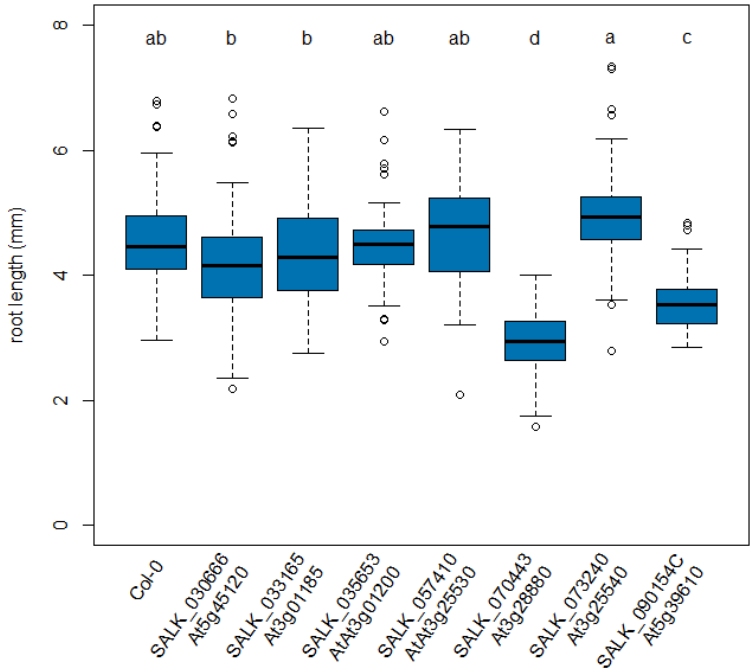

Figure 4. Candidate genes exhibit reduced root growth. T-DNA lines for several candidate genes have been tested for root length. Lines with significantly different root length compared to wild-type are identified (Tukey's post-hoc test).

Supplemental Figures

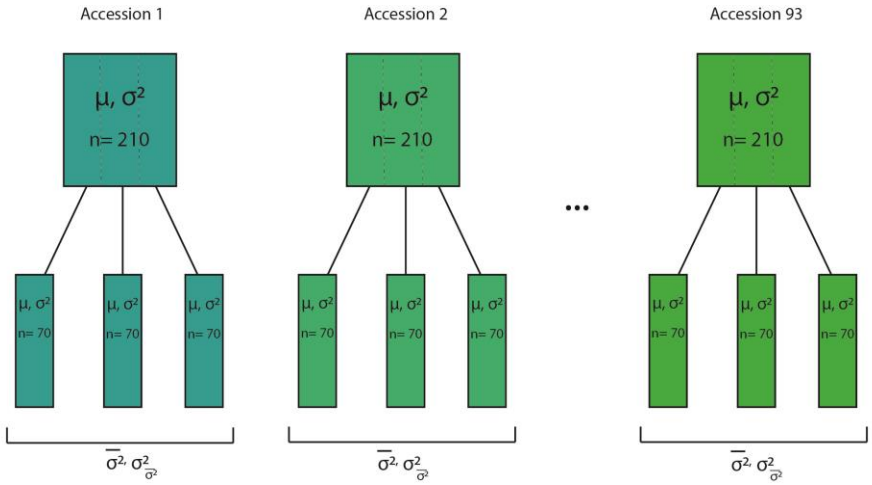

Figure S1. Experimental design to increase accuracy of variance estimates. For 93 accessions, a total of 210 seedlings were grown in three replicates of 70 individuals. For each replicate, the mean and variance was estimated. The mean and variance ($\bar{\sigma}^2, \sigma^2_{\bar{\sigma}^2}$) of the variance across three replicates was also calculated.

accession	stock number
Ag-0	CS22630
An-1	CS22626
Bay-0	CS22633
Bil-5	CS22578
Bor-1	CS22590
Bor-4	CS22591
Br-0	CS22628
Bur-0	CS22656
C24	CS22620
CIBC-17	CS22603
CIBC-5	CS22602
Col-0	CS22625
N13	CS22491
Ct-1	CS22639
CVI-0	CS22614
Eden-1	CS22572
Eden-2	CS22573
Edi-0	CS22657
Ei-2	CS22616
Est-1	CS22629
Fab-2	CS22576
Fab-4	CS22577
Fei-0	CS22645
Ga-0	CS22634
Got-22	CS22609
Got-7	CS22608
Gu-0	CS22617
Gy-0	CS22631
HR-10	CS22597
HR-5	CS22596
Kas-2	CS6751

accession	stock number
Kin-0	CS22654
Knox-10	CS22411
Knox-18	CS22567
Kondara	CS22651
Kz-1	CS22606
Kz-9	CS22607
Ler-1	CS22618
LL-0	CS22650
Löv-1	CS22574
Löv-5	CS22575
Lp2-2	CS22594
Lp2-6	CS22595
Lz-0	CS22615
Mr-0	CS22640
Mrk-0	CS22635
Ms-0	CS22655
Mt-0	CS22642
Mz-0	CS22636
Nd-1	CS22619
NFA-10	CS22599
NFA-8	CS22598
Nok-3	CS22643
ÖMö2-1	CS22584
ÖMö2-3	CS22585
Oy-0	CS22658
Pna-17	CS22570
Pro-0	CS22649
Pu2-23	CS22593
Pu2-7	CS22592
Ra-0	CS22632
Ren-1	CS22610

accession	stock number
Ren-11	CS22611
Rmx-A02	CS22568
Rmx-A180	CS22569
RRS-10	CS22565
RRS-7	CS22564
Se-0	CS22646
Shakdara	CS22652
Sorbo	CS22653
Spr1-2	CS22582
Sq-1	CS22600
Sq-8	CS22601
Tamm-2	CS22604
Tamm-27	CS22605
Ts-1	CS22647
Ts-5	CS22648
Tsu-1	CS22641
Ull2-3	CS22587
Ull2-5	CS22586
Uod-1	CS22612
Uod-7	CS22613
Van-0	CS22627
Vår2-1	CS22580
Vår2-6	CS22581
Wa-1	CS22644
Wei-0	CS22622
Ws-0	CS22623
Ws-2	CS22659
Wt-5	CS22637
Yo-0	CS22624
Zdr-1	CS22588
Zdr-6	CS22589

Table S2. *A. thaliana* accessions phenotyped for GWA.

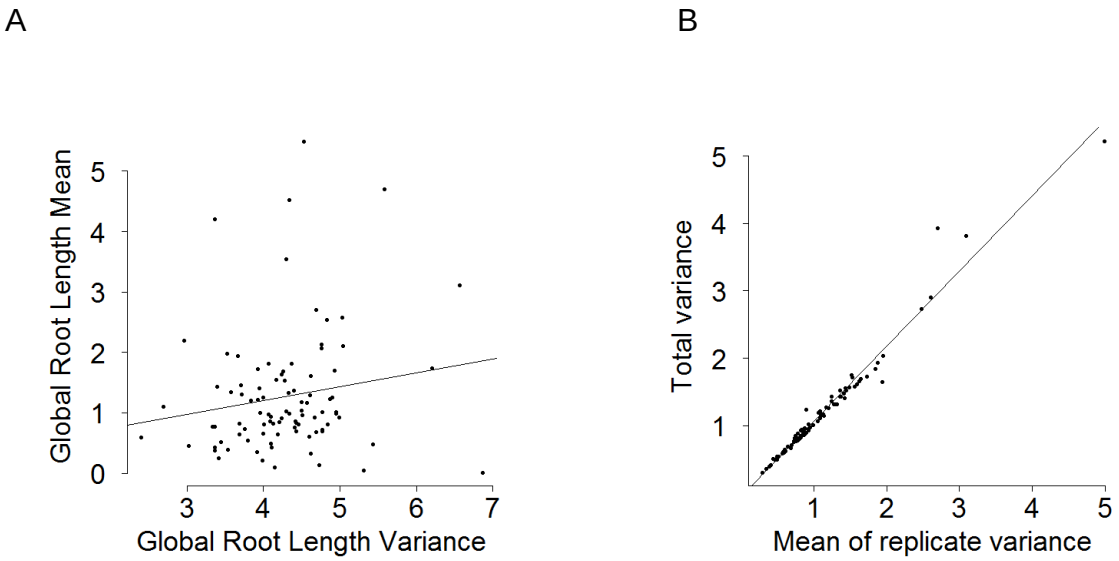

Figure S2. Root length variance varies independently of root mean and can be calculated globally or per replicate. A) There is no significant correlation between root length mean and root length variance ($p = 0.11$, $R^2 = 0.16$, Pearson's). B) There is a strong correlation between the total variance estimated from ~210 individuals and mean variance from three replicates of ~70 individuals ($p < 2.2e\text{-}16$, $R^2 = 0.98$).

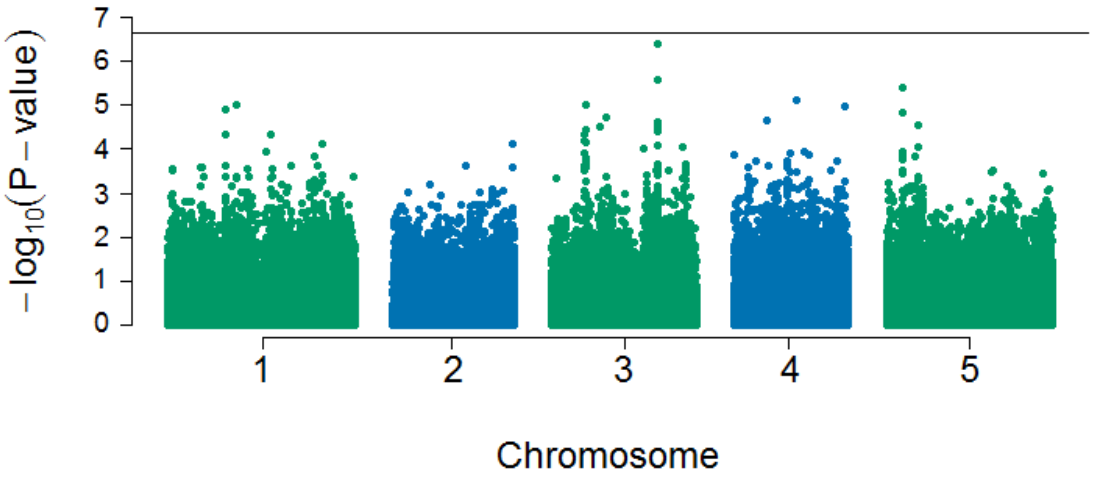

Figure S3. No loci are detected for root length variance using GWAS with an additive model. A GWAS was completed under an additive model. No SNPs reached the Bonferroni threshold.

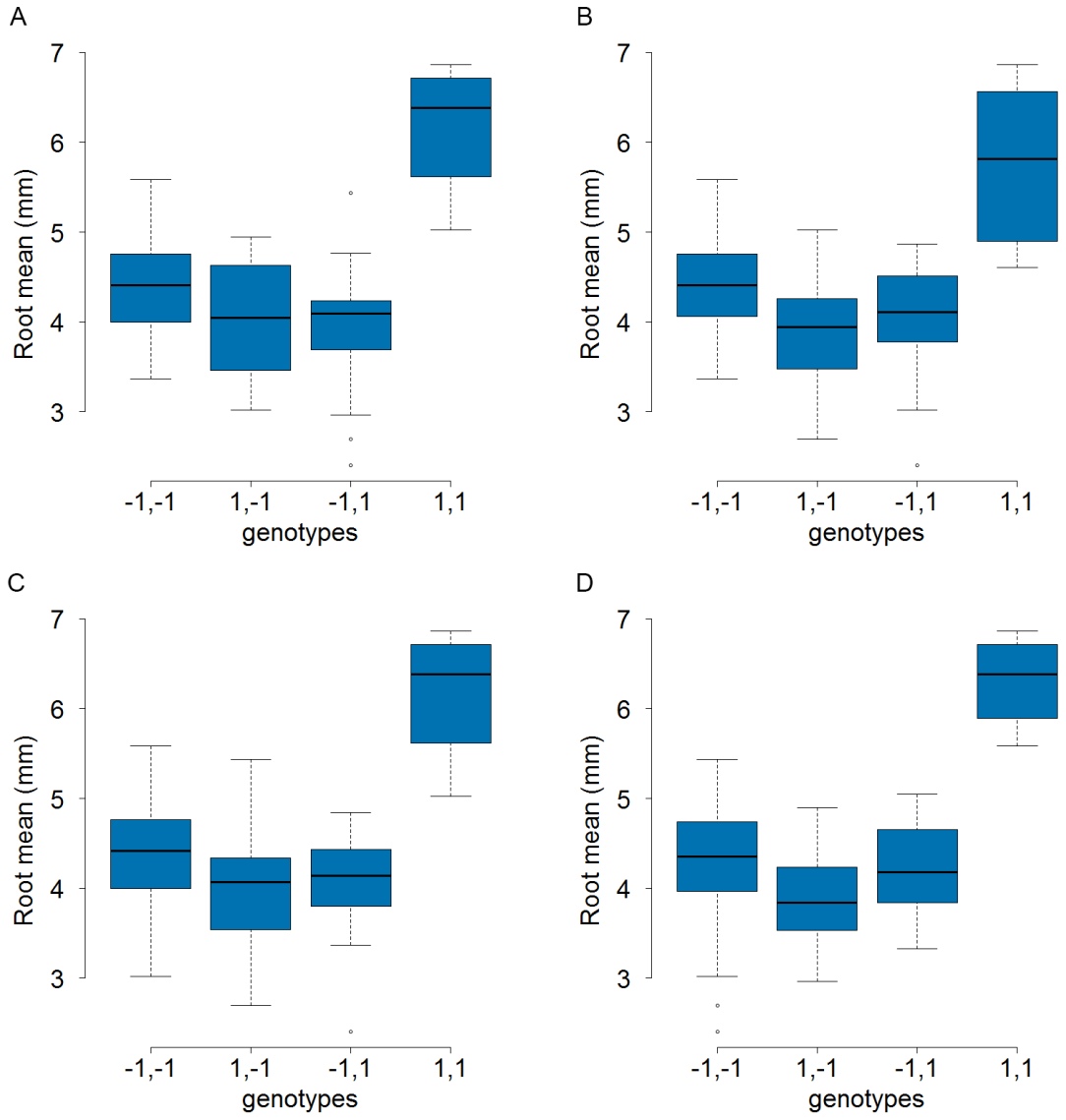
A
B
C
D
Root mean (mm)
7
6
5
4
3
-1,-1
1,-1
-1,1
1,1
genotypes

Figure S4. The same pattern of effects is observed for all significantly interacting loci. A-D) Root mean is displayed for the four pairs of significantly interacting loci. The major allele is indicated by -1 and the minor allele is indicated by 1. The combination of two minor alleles (1,1) always has the largest root length compared to the other three allele combinations, which are more similar to one another. A) Root mean for allele combinations at1_17257526 and 5_15862026 /5_15862525. B) Root mean for allele combinations at 3_66596 and 3_9272294/3_9273674. C) Root mean for allele combination at 3_66596 and 5_18241640. D) Root mean for allele combination at 3_10891195 and 5_1027939.

At3g46700 amiR I	
Oligo I	TGTATTCAAGATAAGGGCTTGGCCCATGATGATCACATTCGTTATCTATTTTTTGGGCCAAGCCATTATCTTGAA
Oligo II	AATGTTCAAGATAATGGCTTGGCCCAAAAAATAGATAACGAATGTGATCATCATGGGCCAAGCCCTTATCTTGAA
At3g46700 amiR II	
Oligo I	TGTATCAAGATAAGGGCTTGGCCGTATGATGATCACATTCGTTATCTATTTTTTACGGCCAAGCACTTATCTTGA
Oligo II	AATGTCAAGATAAGTGCTTGGCCGTAAAAAATAGATAACGAATGTGATCATCATACGGCCAAGCCCTTATCTTGA
At3g46700 amiR III	
Oligo I	TGTATTAGAGACCCGACTGGTCCAAATGATGATCACATTCGTTATCTATTTTTTTTGGACCAGTAGGGTCTCTAA
Oligo II	AATGTTAGAGACCCTACTGGTCCAAAAAAAATAGATAACGAATGTGATCATCATTTGGACCAGTCGGGTCTCTAA

Table S2. Sequences of artificial microRNAs used to target At3g46700.

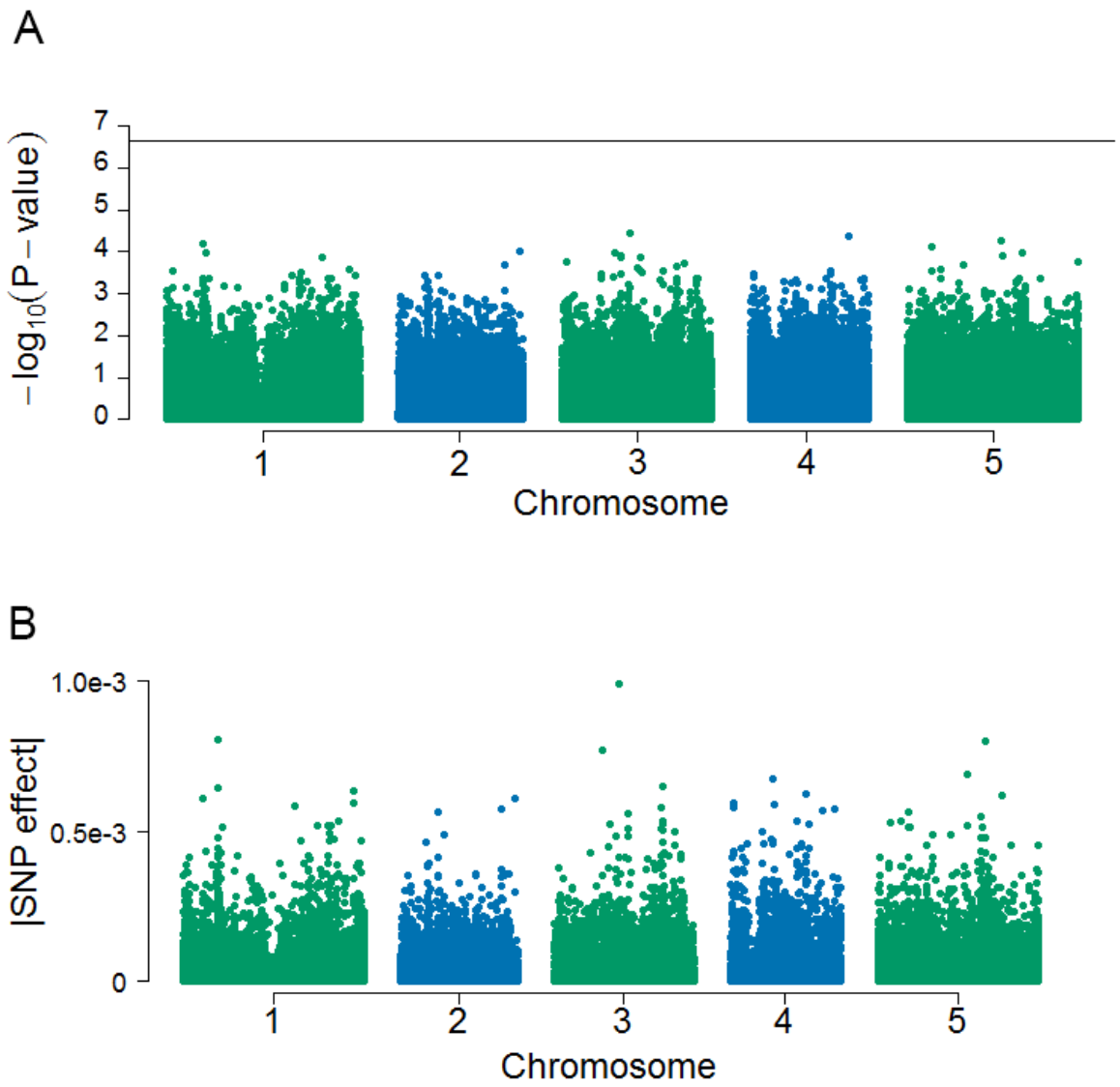

Figure S5. No associations are found with root length using an additive model. A) A GWAS was completed under an additive model. No SNPs reached the Bonferroni threshold. B) Using a whole-genome generalized ridge regression model in which SNPs were random effects, no SNPs reached the Bonferroni threshold.

Gene	T-DNA line	Insertion location
At3g01180	SALK_065639	Exon 2 of 8
At3g01185	SALK_033165	promoter
At3g01200	SALK_035653	Exon 1 of 3
At3g25520	SALK_135037	5'UTR
At3g25530	SALK_057410	Intron 3 of 7
At3g25540	SALK_073240	Exon 1 of 5
At5g45120	SALK_030666	5'UTR
At5g39610	SALK_090154	Exon 3 of 3
At5g39620	SALK_096950	Intergenic
At3g28865	SALK_089352	Exon 1 of 1
At3g28880	SALK_070443	Exon 17 of 19
At5g03840	SALK_142051	5' UTR
At5g03850	SALK_112775	Exon 1 of 1

Table S3. T-DNA lines tested for root phenotypes.

CHAPTER VII

DISCUSSION AND FUTURE DIRECTIONS[1]

I have furthered understanding of Hsp90's ability to buffer genetic variation and developmental noise. Specifically, my data demonstrate that the nucleotide sequences of genes encoding Hsp90 clients exhibit relaxed purifying selection. I also find that Hsp90's interaction with one other gene is enough to regulate developmental noise in the trait in question. Below, I discuss these findings in a broader evolutionary sense by speculating on Hsp90's effects on genome evolution and the genetic architecture underlying developmental noise.

A molecular signature of Hsp90 is found in genomes

The evolutionary capacitor hypothesis states that when Hsp90 function is challenged, novel, adaptive phenotypes are revealed and then selected. An assumption of this hypothesis is that the genetic loci underlying these phenotypes exhibit signals of change in evolutionary rate. My results demonstrate that Hsp90 relaxes selection on genes encoding clients in *S. cerevisiae*, *A. thaliana*, and *H. sapiens* (Chapters II-III, Appendices I-II), consistent with Hsp90's evolutionary effect being highly conserved. Moreover, within *H. sapiens* and *A. thaliana*, I detected a positive effect of Hsp90 interaction on nucleotide diversity at genes that encode clients compared to genes that encode non-clients. Therefore, even at shorter time scales,

[1] Portions of this chapter are adapted from "Molecular mechanisms of robustness in plants," Current Opinion in Plant Biology: 16(1), 2013, pp. 62-69, by J. Lempe, J. Lachowiec, A. M. Sullivan and C. Queitsch.

buffering of genetic variation by Hsp90 can be detected. A lack of a large number of strains may explain why the same was not observed for *S. cerevisiae* (Appendix III).

In order to detect the influence of Hsp90 on evolutionary rate (dN/dS) many factors had to be controlled (Chapter II-III). First and foremost, selection influences dN/dS by acting on gene products and their functions. I attempted to normalize for the functions of genes by focusing on genes that originate from the same phylogenetic group. Expression level (Wall et al. 2005; Pál et al. 2001) and number of protein interactions (Fraser et al. 2002) also shape dN/dS, so correcting for their influence was critical to prevent confounding of Hsp90's effect. The influence of Hsp90 was detected only after comparing related genes and correcting for expression levels and the number of protein interactions. Genes that shared functions but lacked a phylogenetic relationship did not exhibit an enhanced evolutionary rate due to Hsp90, indicating that comparisons among related genes correct for more than function. Using phylogenetically related genes may have also corrected for shared structural features.

If Hsp90 acts as an evolutionary capacitor, we may also expect positive selection at the loci that become beneficial upon loss of buffering. A phenotype that enhances fitness would quickly become fixed in a population along with its causal genotype (Hartl & Clark 2007), and this can occur even when buffering is restored (Rutherford & Lindquist 1998). None of the Hsp90 clients I examined have undergone positive selection, as measured by dN/dS, to a greater degree than non-clients. In fact, none of the proteins, client or not, exhibited positive selection. Only a subset of known Hsp90 clients were inspected in my analyses, and it may be that these particular clients do not participate in functions that are typically positively selected, such as fertility or immunity (Swanson et al. 2001; Nielsen et al. 2005).

dN/dS detects positive selection over divergent lineages (Kryazhimskiy & Plotkin 2008), but Hsp90's role as an evolutionary capacitor instead might be detected at shorter time scales. Signals of selective sweeps, a measure of positive selection in populations, have been detected in almost 10% of the human genome (Akey 2009). Using selective sweeps to identify positive selection in regions that previously harbored cryptic genetic variation was recently suggested and could be applied to the case of Hsp90 (Paaby & Rockman 2014). It is not known if Hsp90 clients are enriched in these regions of selective sweeps, but I expect that some clients occupy those regions based on the large number of known clients and the large extent of positively selected regions. Interestingly, selective sweeps are signaled by regions with low genetic diversity, which is the opposite of the effect seen in Hsp90 kinase clients, demonstrating the complexity of evolution's genomic effects. However looking at positive selection at genes encoding clients is not the whole story. Positive selection at beneficial, Hsp90-revealed loci could also exist.

Models of how Hsp90 perturbation influences phenotypes state that Hsp90 reshapes genetic networks underlying phenotypes as it is titrated away from its ambient-condition clients (Sangster et al. 2004; Lempe et al. 2012). For example, upon stress when competition for interaction with Hsp90 is increased, a transcription factor client may lose function, revealing beneficial variation at its target promoter (Sangster et al. 2004). As synthetic lethal interactions with Hsp90 have been systematically assessed in yeast, the genetic network of Hsp90 is available (McClellan et al. 2007; Zhao et al. 2005). If any of these Hsp90 interactors are positively selected, they may have previously harbored variation revealed by Hsp90 and causal to an adaptive phenotype. The potential sequence of the gene prior to positive selection can be determined through ancestral reconstructions using sequences from closely related populations or

species. Transforming the reconstructed ancestral gene into modern organisms (or even the modern sequence into a related species) and examining its effect on phenotypes, with and without Hsp90 perturbation, may uncover what phenotype(s) changed and how it/they became beneficial. The existence of such a case would demonstrate Hsp90's evolutionary capacitor function.

Hsp90 and functional divergence in genomes

I observed that Hsp90 can buffer genetic variation directly in its clients (Chapter II and III) and that this interaction allows increasing amounts of amino acid sequence flexibility over time (Chapter III). Therefore, Hsp90 flattens fitness landscapes, allowing genes encoding clients to visit a larger genotypic space than genes encoding non-clients or to shift from local to global fitness peaks (Sangster et al. 2004). To empirically demonstrate that Hsp90 can do so, I have designed the following experiment (Figure 6.1): Fuse a promoter that is activated by both the Hsp90 client BES1 and the non-client BZR1 (Chapter II) upstream to the reporter gene HIS3. Integrate this cassette into a background that does not express HIS3. Randomly mutagenize BES1 and BZR1 and separately introduce libraries of mutations into the yeast with the HIS3 reporter. Collect colonies that grow on HIS+ (input) and HIS- media (experimental). Count the number of mutations in functional BES1 and BZR1 through next-generation sequencing in both conditions to assess the number of mutants tested in total (input) and the number of tolerated mutations (experimental). My hypothesis is that BES1 will tolerate more mutations than BZR1 due to its interaction with Hsp90. I can repeat the experiment using media containing an Hsp90 inhibitor to demonstrate that the mutations are Hsp90 dependent. Hsp90's buffering ability is

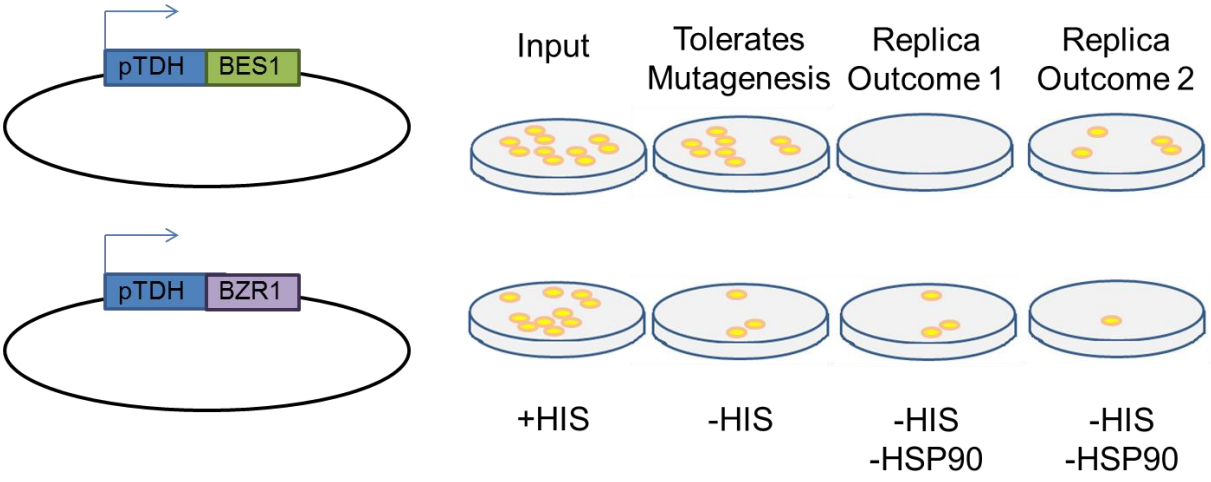

Figure 6.1. Assaying Hsp90 dependence in randomly mutagenized libraries of *BES1* and *BZR1*. A plasmid library of mutagenized *BES1* or *BZR1* under the control of a yeast promoter are transformed into yeast carrying a cassette in which HIS3 is regulated by a promoter similarly bound and activated by functional BES1 or BZR1. In the first column of plates is the total number of mutagenized *BES1* or *BZR1* variants. In the second column of plates, I hypothesize that more variants of *BES1* will tolerate mutagenesis than *BZR1* because of BES1's client status. In column two and three, the colonies are replica plated onto media containing an Hsp90 inhibitor. In Outcome 1, all of the BES1 variants are Hsp90-dependent and do not grow in the presence of the Hsp90 inhibitor, whereas the BZR1 variants are all Hsp90 independent and grow in the presence of the Hsp90 inhibitor. In Outcome 2, some variants of BES1 become independent of Hsp90 and survive Hsp90 inhibition, whereas some variants of BZR1 become Hsp90-dependent and die in the presence of the Hsp90 inhibitor.

cleanly evaluated with this system, because the experiment uses two nearly identical genes with differential client status.

Hsp90 is implicated in the functional divergence of gene families because it permits proteins to visit a larger functional landscape. I speculate that the frequent interaction with Hsp90 throughout evolution contributed to both the number and the range in function of the kinases. Specifically, Hsp90 could decrease the rate of gene death to promote the size of gene families. Normally, sequences of duplicated genes drift and decay due to the presence of functionally redundant copies, leading to the loss of duplicated genes from the genome (Conant & Wolfe 2008).

If the initial mutations that lead to the decay of a duplicated gene are buffered by Hsp90, the gene's lifespan may increase. With increased lifespans and the ability to tolerate more mutations, Hsp90 clients can explore more functional space (Conant & Wolfe 2008). In doing so, a gene duplicate encoding a client may acquire a useful, non-redundant function that preserves the gene in the genome. I propose that Hsp90 lengthens the lifespan of genes by shielding genes from death and promoting functional divergence, thus contributing to genome evolution.

Duplicated genes also diverge in function through changes in the degree (Duarte et al. 2006), timing, and/or location (Liu et al. 2011) of transcription. This could also happen via Hsp90. Interaction with Hsp90 *de facto* alters protein expression. When Hsp90 is challenged by changing conditions, there is thought to be an increase in the number of proteins requiring Hsp90 (Sangster & Queitsch 2005). If Hsp90 is titrated away from ambient-condition clients, those clients may no longer function, and selection for their duplicate to be maintained may increase.

Upon stress, Hsp90-dependent proteins may differ from tissue to tissue, cell to cell, and even organelle to organelle, creating complicated expression patterns of clients compared to their non-client relatives. If Hsp90 regulation of protein expression is beneficial and maintains genes in duplicate, one may expect gene families with Hsp90 clients to be larger in size than those without clients. Once more gene families are systematically tested for their interaction with Hsp90, the number of Hsp90 interactions can be tested for correlation to gene family size.

What makes an Hsp90 client?

The identifying features of Hsp90 client status have been elusive, but evidence points to the stability of proteins as the defining attribute (Taipale et al. 2012). Using the experimental system described in Figure 6.1, a mutational scan of amino acids that imbue client status can be completed. By measuring the dependency of BES1 and BZR1 mutants on Hsp90, amino acids that are important to client status could be identified. BES1 and BZR1 only differ by 30 amino acids, and it may be that some small number of mutations can switch BES1 and BZR1's client statuses. Furthermore, the stability of these mutants could be measured by thermal denaturation to see if it correlates to Hsp90 dependency as expected.

Differences in Hsp90 dependency has been observed for client proteins. In addition to the continuum of Hsp90 interaction with kinases (Taipale et al. 2012), the mammalian steroid hormone receptors also have a varying dependence on Hsp90 (Pratt 1990) (Appendix II). The continuum of dependence can arise from differences in the length of time each protein interacts with Hsp90 or from differences in the proportion of the protein population requiring Hsp90 or a combination of the two. It also may be that clients repeatedly cycle through interaction with

Hsp90, as has been observed with Hsp70 (Kampinga & Craig 2010). The number of cycles to reach a mature protein fold may be what defines this Hsp90 client interaction continuum. Why this Hsp90 dependency varies will be revealed as we learn more about the structure and mechanism of the interaction between Hsp90 and its substrates.

A single gene interaction with Hsp90 regulates robustness to developmental noise

Hsp90's buffering of genetic variation and developmental noise is thought to arise from its position as a hub in genetic networks. At least one Hsp90 client (Chapter II) and possibly three (Chapter IV) are encoded by the *BEH* family, suggesting that Hsp90 is tightly connected to the family. Interestingly, *BEH4* was a strong regulator of robustness in hypocotyl length.

As BES1 is an Hsp90 client (Chapter II), I suspect that BEH4 is as well, since *beh4-1* and *bes1-2* shared a similar hypocotyl length mean response to Hsp90 inhibition. Response to Hsp90 inhibition alone does not designate a client (Taipale et al. 2010), but it is an indicator. The response to Hsp90 inhibition in *beh4-1* hypocotyl length robustness was surprising. Unlike *bes1-2*, *beh4-1* exhibited no further loss in robustness when Hsp90 activity was inhibited, demonstrating an interaction between *BEH4* and Hsp90 in robustness. Combining these findings suggests that BEH4 is a client of Hsp90 that maintains variance in hypocotyl length, but upon Hsp90 inhibition, BEH4 loses function, explaining the increased in variance.

Evidence suggests that robustness in genetic networks is an emergent trait (Siegal & Bergman 2002). Hsp90's position as a hub in genetic networks is thought to explain why Hsp90 can buffer genetic variation and developmental noise to maintain robust phenotypes (Sangster et al. 2004). Current models suggest that when Hsp90 is inhibited, the activity of many other genes

regulating the trait of interest have altered functionality and result in an increase in developmental noise (Sangster et al. 2004). My work on the *BEH* family suggested otherwise. For the trait hypocotyl length in the conditions tested, robustness due to Hsp90 appears to be provided through its interaction with one other gene, *BEH4*, not many. In yeast, variation in cellular morphology due to Hsp90 was also traced to the activity of a single gene (Hsieh et al. 2013). Hsp90's position as a hub in genetic networks may instead explain the broad reach of Hsp90: its effect on the variation of a large number of traits (Todd A Sangster et al. 2008; T Sangster et al. 2008).

Identifying genes regulating developmental noise levels

In the case of *BEH4*, I was not searching for a gene that regulated developmental noise; its identification was by chance. I also took a deliberate approach to identify novel regulators of developmental noise using the densely genotyped collection of *A. thaliana* accessions with the knowledge that level of developmental noise is a heritable, quantitative trait (Hall et al. 2007; Todd A Sangster et al. 2008; Lempe et al. 2012). GWAS and subsequent linkage analysis identified the gene At3g46700 as harboring genetic variation regulating developmental noise (Chapter V). This gene shares homology with five other genes in the immediate vicinity, so I suspect that its ability to regulate noise arises from its redundant function with its close relatives. Similar to other studies (Ordas et al. 2008; Todd A Sangster et al. 2008; Jimenez-Gomez et al. 2011), we find that this gene is not mapped for trait mean. In fact, the genetic architecture underlying trait mean was entirely different. For variance, there was an additive contribution to heritability; whereas for mean, a larger contribution was predicted to be epistatic than additive.

Future efforts in regulating developmental noise

While GWAS could identify variation at one locus that interacts with developmental noise in one trait, it did not detect multiple loci, nor did it identify any loci regulating developmental noise in the other trait I examined (Appendix IV). What is the best what to characterize regulators of developmental noise going forward? Theory (Wagner 2005; Masel & Siegal 2009) and some empirical evidence (Lehner 2010; Levy & Siegal 2008) indicate that buffers of genetic variation, developmental noise, and environmental perturbations are the same. Therefore, it seems reasonable to study any buffer for its ability to buffer developmental noise. For example, yeast and worm studies strongly implicate chromatin modifiers in robustness (Levy & Siegal 2008; Lehner et al. 2006). Also, maize mutants in RNA-directed DNA methylation show stochastic developmental defects (Hollick 2010; Parkinson et al. 2007) (Fig. 1.1c). Further, defects in ribosome function result in highly pleiotropic developmental defects in humans (Boulon et al. 2010; Freed et al. 2010) and maize (Degenhardt & Bonham-Smith 2008), suggesting that genes involved in ribosome biogenesis, rRNA processing, and RNA splicing may also maintain robustness.

Agriculture is one field of many that benefits from understanding properties of robustness. Uniformity is the term used for robustness in crops, and it is highly desirable. In maize, uniform plant height facilitates harvest, maximizing the amount of ears harvested, and in animal breeding, constant litter size helps minimize resource use (SanCristobal-Gaudy et al. 2001). In fact, seed size uniformity is one of the signs anthropologists use to identify the domestication of plants (Perry 2002). The primary method to increase uniformity among crops is to reduce genetic variation. For example, genetic variation in maize is eliminated by only

planting hybrids for harvest. Furthermore, most studies of crop robustness focus on robustness to the environment such as variation in water availability and soil conditions (Paterson et al. 2003; Makumburage & Stapleton 2011). However, breeding for reduced developmental noise may also aid in these efforts. Identifying and manipulating molecular mechanisms that provide robustness to developmental noise may circumvent hybrid monoculture, in which the whole population, rather than a subset if individuals, is susceptible to stresses such as disease and pests.

One of the major draw-backs for current robustness studies is the large sample sizes required for population-based robustness measures. Thus far, individual-based robustness measures remain less amenable to high-throughput analysis (Appendix VI) (Queitsch et al. 2012). We speculate that comparing organisms perturbed in functionally distinct robustness master regulators may reveal shared molecular features such as specific changes in gene expression, methylation or nucleolar function. These shared features could be leveraged as molecular robustness markers that are applicable to individuals and large populations. Molecular robustness markers would revolutionize the study of robustness in non-model organisms, including humans, and allow us to explore the role of robustness in evolutionary processes and disease susceptibility.

Final thoughts on Hsp90 and evolution

To conclude, I will summarize my thoughts on Hsp90 and evolvability of phenotypes. I am of the mind that the evolutionary effects of Hsp90 are entirely due to its function as a protein chaperone. Second order selection for Hsp90 function as a buffer of genetic variation or developmental noise is unlikely given the rarity of the environmental events that would reveal

such variation. Based on this, I doubt that Hsp90 releases cryptic genetic variation that is adaptive to the specific environments that regulate Hsp90 function.

However, there is clear evidence that both of the buffering abilities of Hsp90 contribute to the final distribution of phenotypes. An environmental trigger that causes the loss of the Hsp90 buffer would not only reveal cryptic genetic variation, but also simultaneously increase developmental noise. Revealed cryptic genetic variants can become selected and independent of the inhibition of the Hsp90 (Rutherford & Lindquist 1998), whereas the unbuffering of developmental noise requires a constant loss of Hsp90 function and selection for this loss of Hsp90 function. Therefore, the evolutionary contribution of Hsp90 in regulating development noise may not be reflected in modern phenotypes. This lack of long-term phenotypic effect does not mean that developmental noise buffering is evolutionarily irrelevant. For example, an increase in developmental noise may be critical to the survival of a population, especially a population with low genetic diversity, to short-term environmental assault that affects Hsp90 function and shifts the fitness landscape (Kærn et al. 2005). The future in understanding Hsp90 as a buffer of genetic variation and developmental noise includes knowledge of both its short-term and long-term evolutionary effects.

REFERENCES

1. Adzhubei IA et al. 2010. A method and server for predicting damaging missense mutations. Nat Meth. 7:248–249. doi: http://www.nature.com/nmeth/journal/v7/n4/suppinfo/nmeth0410-248_S1.html.

2. Aeong Oh S et al. 1997. Identification of three genetic loci controlling leaf senescence in Arabidopsis thaliana. Plant J. 12:527–535. doi: 10.1046/j.1365-313X.1997.00489.x.

3. Akey JM. 2009. Constructing genomic maps of positive selection in humans: Where do we go from here? Genome Res. 19:711–722.

4. Albrecht C, Russinova E, Kemmerling B, Kwaaitaal M, de Vries SC. 2008. Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASE Proteins Serve Brassinosteroid-Dependent and -Independent Signaling Pathways. Plant Physiol. 148:611–619. doi: 10.1104/pp.108.123216.

5. Allan WL, Simpson JP, Clark SM, Shelp BJ. 2008. γ-Hydroxybutyrate accumulation in Arabidopsis and tobacco plants is a general response to abiotic stress: putative regulation by redox balance and glyoxylate reductase isoforms. J. Exp. Bot. 59:2555–2564. doi: 10.1093/jxb/ern122.

6. Ansel J et al. 2008. Cell-to-cell stochastic variation in gene expression is a complex genetic trait. PLoS Genet. 4:e1000049.

7. Asami T et al. 2000. Characterization of brassinazole, a triazole-type brassinosteroid biosynthesis inhibitor. Plant Physiol. 123:93–100. http://www.ncbi.nlm.nih.gov/pubmed/10806228.

8. Atwell S et al. 2010. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature. 465:627–631. doi: 10.1038/nature08800 nature08800 [pii].

9. Aulchenko YS, Ripke S, Isaacs A, Van Duijn CM. 2007. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 23:1294–1296.

10. Aury JM et al. 2006. Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. Nature. 444:171–178. doi: nature05230 [pii] 10.1038/nature05230.

11. Bai M-Y et al. 2012. Brassinosteroid, gibberellin and phytochrome impinge on a common transcription module in Arabidopsis. Nat. Cell Biol. 14:810–817.

12. Beers EP, Jones AM, Dickerman AW. 2004. The S8 serine, C1A cysteine and A1 aspartic protease families in Arabidopsis. Phytochemistry. 65:43–58. http://www.sciencedirect.com/science/article/pii/S003194220300565X (Accessed December 5, 2013).

13. Bergman A, Siegal ML. 2003. Evolutionary capacitance as a general feature of complex gene networks. Nature. 424:549–552. doi: 10.1038/nature01765.

14. Bielawski JP, Yang. 2003. Maximum likelihood methods for detecting adaptive evolution after gene duplication. J. Struct. Funct. Genomics. 3:201–212. http://www.ncbi.nlm.nih.gov/pubmed/12836699.

15. Blanc G, Hokamp K, Wolfe KH. 2003. A Recent Polyploidy Superimposed on Older Large-Scale Duplications in the Arabidopsis Genome . Genome Res. 13:137–144. doi: 10.1101/gr.751803.

16. Bloom JD, Adami C. 2003. Apparent dependence of protein evolutionary rate on number of interactions is linked to biases in protein–protein interactions data sets. BMC Evol. Biol. 3:21. doi: 10.1186/1471-2148-3-21.

17. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. 2006. Protein stability promotes evolvability. Proc. Natl. Acad. Sci. U. S. A. 103:5869–5874. doi: 10.1073/pnas.0510098103.

18. Bogumil D, Alvarez-Ponce D, Landan G, McInerney JO, Dagan T. 2014. Integration of Two Ancestral Chaperone Systems into One: The Evolution of Eukaryotic Molecular Chaperones in Light of Eukaryogenesis. Mol. Biol. Evol. 31:410–418. doi: 10.1093/molbev/mst212.

19. Bogumil D, Dagan T. 2010. Chaperonin-dependent accelerated substitution rates in prokaryotes. Genome Biol. Evol. 2:602–608. doi: 10.1093/gbe/evq044.

20. Bogumil D, Landan G, Ilhan J, Dagan T. 2012. Chaperones divide yeast proteins into classes of expression level and evolutionary rate. Genome Biol. Evol. 4:618–25. doi: 10.1093/gbe/evs025.

21. Borkovich KA, Farrelly FW, Finkelstein DB, Taulien J, Lindquist SL. 1989. hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. Mol. Cell. Biol. 9:3919–3930.

22. Boulon S, Westman BJ, Hutten S, Boisvert F-M, Lamond AI. 2010. The nucleolus under stress. Mol. Cell. 40:216–227.

23. Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature. 422:433–438.

24. Brady SM et al. 2011. A stele-enriched gene regulatory network in the Arabidopsis root. Mol. Syst. Biol. 7.

25. Breitkreuz KE et al. 2003. A novel gamma-hydroxybutyrate dehydrogenase: identification and expression of an Arabidopsis cDNA and potential role under oxygen deficiency. J. Biol. Chem. 278:41552–6. doi: 10.1074/jbc.M305717200.

26. Brodersen P, Voinnet O. 2006. The diversity of RNA silencing pathways in plants. TRENDS Genet. 22:268–280.

27. Burga A, Casanueva MO, Lehner B. 2011. Predicting mutation outcome from early stochastic variation in genetic interaction partners. Nature. 480:250–253. doi: http://www.nature.com/nature/journal/v480/n7376/abs/nature10665.html#supplementary-information.

28. Cao J et al. 2011. Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat. Genet. 43:956–963.

29. Carbonell A et al. 2014. New Generation of Artificial MicroRNA and Synthetic Trans-Acting Small Interfering RNA Vectors for Efficient Gene Silencing in Arabidopsis. Plant Physiol. 165:15–29.

30. Carlborg Ö, Jacobsson L, Åhgren P, Siegel P, Andersson L. 2006. Epistasis and the release of genetic variation during long-term selection. Nat. Genet. 38:418–420.

31. Casanueva MO, Burga A, Lehner B. 2012. Fitness trade-offs and environmentally induced mutation buffering in isogenic C. elegans. Science (80-.). 335:82–85. doi: 10.1126/science.1213491.

32. Castle JC et al. 2010. Digital genome-wide ncRNA expression, including SnoRNAs, across 11 human tissues using polyA-neutral amplification. PLoS One. 5:e11779. doi: 10.1371/journal.pone.0011779.

33. Cederholm HM, Iyer-Pascuzzi AS, Benfey PN. 2012. Patterning the primary root in Arabidopsis. Wiley Interdiscip. Rev. Dev. Biol. 1:675–691.

34. Chen B, Wagner A. 2012. Hsp90 is important for fecundity, longevity, and buffering of cryptic deleterious variation in wild fly populations. BMC Evol. Biol. 12:25.

35. Chen M, Markham JE, Dietrich CR, Jaworski JG, Cahoon EB. 2008. Sphingolipid long-chain base hydroxylation is important for growth and regulation of sphingolipid content and composition in Arabidopsis. Plant Cell Online. 20:1862–1878.

36. Chinwalla AT. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature. 420:520–562. doi: http://www.nature.com/nature/journal/v420/n6915/suppinfo/nature01262_S1.html.

37. Chitwood DH et al. 2009. Pattern formation via small RNA mobility. Genes Dev. 23:549–554.

38. Chory J. 1999. Brassinosteroids and the control of plant stature. In: FASEB JOURNAL.Vol. 13 FEDERATION AMER SOC EXP BIOL 9650 ROCKVILLE PIKE, BETHESDA, MD 20814-3998 USA pp. A1339–A1339.

39. Cid C, Garcia-Descalzo L, Casado-Lafuente V, Amils R, Aguilera A. 2010. Proteomic analysis of the response of an acidophilic strain of Chlamydomonas sp.(Chlorophyta) to natural metal-rich water. Proteomics. 10:2026–2036.

40. Citri A et al. 2006. Hsp90 recognizes a common surface on client kinases. J Biol Chem. 281:14361–14369. doi: M512613200 [pii] 10.1074/jbc.M512613200.

41. Clare DK et al. 2012. ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. Cell. 149:113–23. doi: 10.1016/j.cell.2012.02.047.

42. Clark CB. 2009. The role of Hsp90 in PC-12 cell survival. University of Louisville.

43. Clarke GM. 1998. The genetic basis of developmental stability. V. Inter-and intra-individual character variation. Heredity (Edinb). 80:562–567.

44. Clough SJ, Bent AF. 1998. Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana. plant J. 16:735–743.

45. Clouse S, Hall A, Langford M, McMorris T, Baker M. 1993. Physiological and molecular effects of brassinosteroids on <i>Arabidopsis thaliana</i&gt. J. Plant Growth Regul. 12:61–66. doi: 10.1007/bf00193234.

46. Clouse SD. 2002. Brassinosteroids. Arabidopsis Book. 9:e0151. doi: 10.1199/tab.0151.

47. Clouse SD, Zurek DM, McMorris TC, Baker ME. 1992. Effect of brassinolide on gene expression in elongating soybean epicotyls. Plant Physiol. 100:1377–1383.

48. Coluccio MP, Sanchez SE, Kasulin L, Yanovsky MJ, Botto JF. 2011. Genetic mapping of natural variation in a shade avoidance response: ELF3 is the candidate gene for a QTL in hypocotyl growth regulation. J. Exp. Bot. 62:167–176.

49. Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet. 9:938–950. doi: nrg2482 [pii] 10.1038/nrg2482.

50. Cooper GM et al. 2005. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 15:901–913. doi: 10.1101/gr.3577405.

51. Cowen LE, Lindquist SL. 2005. Hsp90 potentiates the rapid evolution of new traits: drug resistance in diverse fungi. Science. 309:2185–9. doi: 10.1126/science.1118370.

52. Darwin C. 1859. On the origin of the species by natural selection.

53. Dean EJ, Davis JC, Davis RW, Petrov DA. 2008. Pervasive and persistent redundancy among duplicated genes in yeast. PLoS Genet. 4:e1000113. doi: 10.1371/journal.pgen.1000113.

54. Debat V, Milton CC, Rutherford S, Klingenberg CP, Hoffmann AA. 2006. Hsp90 and the quantitative variation of wing shape in Drosophila melanogaster. Evolution (N. Y). 60:2529–2538.

55. Degenhardt RF, Bonham-Smith PC. 2008. Arabidopsis ribosomal proteins RPL23aA and RPL23aB are differentially targeted to the nucleolus and are disparately required for normal development. Plant Physiol. 147:128–142.

56. DeLuna A et al. 2008. Exposing the fitness contribution of duplicated genes. Nat Genet. 40:676–681. doi: ng.123 [pii] 10.1038/ng.123.

57. DePristo MA, Weinreich DM, Hartl DL. 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. Nat. Rev. Genet. 6:678–687. doi: 10.1038/nrg1672.

58. Doyle MR et al. 2002. The ELF4 gene controls circadian rhythms and flowering time in Arabidopsis thaliana. Nature. 419:74–77.

59. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed proteins evolve slowly. Proc. Natl. Acad. Sci. U. S. A. 102:14338–14343. doi: 10.1073/pnas.0504070102.

60. Duarte JM et al. 2006. Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis. Mol. Biol. Evol. 23:469–78. doi: 10.1093/molbev/msj051.

61. Durek P et al. 2010. PhosPhAt: the Arabidopsis thaliana phosphorylation site database. An update. Nucleic Acids Res. 38:D828–34. doi: 10.1093/nar/gkp810.

62. Dworkin I, Palsson A, Birdsall K, Gibson G. 2003. Evidence that Egfr Contributes to Cryptic Genetic Variation for Photoreceptor Determination in Natural Populations of Drosophila melanogaster. Curr. Biol. 13:1888–1893. doi: 10.1016/j.cub.2003.10.001.

63. Edelman GM, Gally JA. 2001. Degeneracy and complexity in biological systems. Proc. Natl. Acad. Sci. 98:13763–13768.

64. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–7. doi: 10.1093/nar/gkh340.

65. Ehrlich ES et al. 2009. Regulation of Hsp90 client proteins by a Cullin5-RING E3 ubiquitin ligase. Proc. Natl. Acad. Sci. 106:20330–20335. doi: 10.1073/pnas.0810571106.

66. Eisen JA. 1998. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Res. 8:163–167.

67. Engel SR et al. 2013. The Reference Genome Sequence of Saccharomyces cerevisiae: Then and Now. G3 Genes| Genomes| Genet. g3–113.

68. Felsenstein J. 1985. Phylogenies and the Comparative Method. Am. Nat. 125:1–15. http://www.indiana.edu/~kettlab/A501/Felsenstein1985.pdf.

69. Fernández A, Lynch M. 2011. Non-adaptive origins of interactome complexity. Nature. 474:502–505. doi: 10.1038/nature09992.

70. Finn RD et al. 2006. Pfam: clans, web tools and services. Nucleic Acids Res. 34:D247–D251.

71. Folta A et al. 2014. Over-expression of Arabidopsis AtCHR23 chromatin remodeling ATPase results in increased variability of growth and gene expression. BMC Plant Biol. 14:76.

72. Force A et al. 1999. Preservation of Duplicate Genes by Complementary, Degenerative Mutations. Genetics. 151:1531–1545. http://www.genetics.org/content/151/4/1531.abstract?ijkey=c2226b8d6057bbc72a583a60f7d92f975ec4f6fe&keytype2=tf_ipsecsha (Accessed April 24, 2014).

73. Fraser D, Kærn M. 2009. A chance at survival: gene expression noise and phenotypic diversification strategies. Mol. Microbiol. 71:1333–1340.

74. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary rate in the protein interaction network. Science. 296:750–2. doi: 10.1126/science.1068696.

75. Freed EF, Bleichert F, Dutca LM, Baserga SJ. 2010. When ribosomes go bad: diseases of ribosome biogenesis. Mol. Biosyst. 6:481–493.

76. Fu J et al. 2009. System-wide molecular evidence for phenotypic buffering in Arabidopsis. Nat Genet. 41:166–167. doi: 10.1038/ng.308 ng.308 [pii].

77. Fujikura U, Horiguchi G, Ponce MR, Micol JL, Tsukaya H. 2009. Coordination of cell proliferation and cell expansion mediated by ribosome-related processes in the leaves of Arabidopsis thaliana. Plant J. 59:499–508.

78. Fujiwara K, Ishihama Y, Nakahigashi K, Soga T, Taguchi H. 2010. A systematic survey of in vivo obligate chaperonin-dependent substrates. EMBO J. 29:1552–1564.

79. Gallego-Bartolomé J et al. 2012. Molecular mechanism for the interaction between gibberellin and brassinosteroid signaling pathways in Arabidopsis. Proc. Natl. Acad. Sci. 109:13446–13451.

80. Gampala SS et al. 2007. An Essential Role for 14-3-3 Proteins in Brassinosteroid Signal Transduction in Arabidopsis. Dev. Cell. 13:177–189. doi: 10.1016/j.devcel.2007.06.009.

81. Gangaraju VK et al. 2011. Drosophila Piwi functions in Hsp90-mediated suppression of phenotypic variation. Nat. Genet. 43:153–158. doi: 10.1038/ng.743.

82. Garland T, Harvey PH, Ives AR. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. Syst. Biol. 41:18–32.

83. Gasch AP et al. 2000. Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell. 11:4241–4257. http://www.ncbi.nlm.nih.gov/pubmed/11102521.

84. Gazzani S, Gendall AR, Lister C, Dean C. 2003. Analysis of the molecular basis of flowering time variation in Arabidopsis accessions. Plant Physiol. 132:1107–1114.

85. Geiler-Samerotte K et al. 2013. The details in the distributions: why and how to study phenotypic variability. Curr. Opin. Biotechnol. 24:752–9. doi: 10.1016/j.copbio.2013.03.010.

86. Gerspacher C et al. 2009. The effect of cadmium on brain cells in culture. Int. J. Mol. Med. 24:311.

87. Giannini A, Bijlmakers M-J. 2004. Regulation of the Src family kinase Lck by Hsp90 and ubiquitination. Mol. Cell. Biol. 24:5667–76. doi: 10.1128/MCB.24.13.5667-5676.2004.

88. Gibbs RA et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature. 428:493–521. doi: http://www.nature.com/nature/journal/v428/n6982/suppinfo/nature02426_S1.html.

89. Gibson G. 2012. Rare and common variants: twenty arguments. Nat. Rev. Genet. 13:135–145. doi: 10.1038/nrg3118.

90. Gibson G, Dworkin I. 2004. Uncovering cryptic genetic variation. Nat Rev Genet. 5:681–690. doi: 10.1038/nrg1426 nrg1426 [pii].

91. Goodstein DM et al. 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 40:D1178–D1186.

92. Gu Z et al. 2003. Role of duplicate genes in genetic robustness against null mutations. Nature. 421:63–66.

93. Guan Y, Dunham MJ, Troyanskaya OG. 2007. Functional analysis of gene duplications in Saccharomyces cerevisiae. Genetics. 175:933–943. doi: genetics.106.064329 [pii] 10.1534/genetics.106.064329.

94. Guindon S et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307–321.

95. Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704. doi: 10.1080/10635150390235520.

96. Guo et al. 2009. Three related receptor-like kinases are required for optimal cell elongation in Arabidopsis thaliana. Proc. Natl. Acad. Sci. 106:7648–7653. doi: 10.1073/pnas.0812346106.

97. Gutiérrez J, Maere S. 2014. Modeling the evolution of molecular systems from a mechanistic perspective. Trends Plant Sci. 19:292–303.

98. Hahn MW. 2009. Distinguishing among evolutionary models for the maintenance of gene duplicates. J. Hered. 100:605–617. doi: 10.1093/jhered/esp047.

99. Hall MC, Dworkin I, Ungerer MC, Purugganan M. 2007. Genetics of microenvironmental canalization in Arabidopsis thaliana. Proc Natl Acad Sci U S A. 104:13717–13722. doi: 0701936104 [pii] 10.1073/pnas.0701936104.

100. Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. 2008. GEIGER: investigating evolutionary radiations. Bioinformatics. 24:129–131.

101. Harrison R, Papp B, Pál C, Oliver SG, Delneri D. 2007. Plasticity of genetic interactions in metabolic networks of yeast. Proc. Natl. Acad. Sci. 104:2307–2312.

102. Hartl DL, Clark AG. 2007. Principles of Population Genetics. Sinauer Associates, Incorporated http://books.google.com/books?id=SB1vQgAACAAJ.

103. Hartl FU, Bracher A, Hayer-Hartl M. 2011. Molecular chaperones in protein folding and proteostasis. Nature. 475:324–32. doi: 10.1038/nature10317.

104. He et al. 2005. BZR1 is a transcriptional repressor with dual roles in brassinosteroid homeostasis and growth responses. Science (80-.). 307:1634–1638. doi: 10.1126/science.1107580.

105. Heazlewood JL et al. 2008. PhosPhAt: a database of phosphorylation sites in Arabidopsis thaliana and a plant-specific phosphorylation site predictor. Nucleic Acids Res. 36:D1015–21. doi: 10.1093/nar/gkm812.

106. Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics. 22:2971–2. doi: 10.1093/bioinformatics/btl505.

107. Helariutta Y et al. 2000. The *SHORT-ROOT* Gene Controls Radial Patterning of the *Arabidopsis* Root through Radial Signaling. Cell. 101:555–567.

108. Hemani G et al. 2014. Detection and replication of epistasis influencing transcription in humans. Nature.

109. Hernandez RD et al. 2011. Classic selective sweeps were rare in recent human evolution. Science (80-.). 331:920–924.

110. Hill WG, Goddard ME, Visscher PM. 2008. Data and theory point to mainly additive genetic variance for complex traits. Mackay, TFC, editor. PLoS Genet. 4:e1000008. doi: 10.1371/journal.pgen.1000008.

111. Le Hir H, Nott A, Moore MJ. 2003. How introns influence and enhance eukaryotic gene expression. Trends Biochem. Sci. 28:215–220. doi: 10.1016/s0968-0004(03)00052-5.

112. Hollick JB. 2010. Paramutation and development. Annu. Rev. Cell Dev. Biol. 26:557–579.

113. Holmes-Davis R et al. 2005. Proteome mapping of mature pollen of Arabidopsis thaliana. Proteomics. 5:4864–4884. doi: 10.1002/pmic.200402011.

114. Hornstein E, Shomron N. 2006. Canalization of development by microRNAs. Nat. Genet. 38:S20–S24.

115. Hruz T et al. 2008. Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. Adv. Bioinformatics. 2008.

116. Hsieh Y-Y, Hung P-H, Leu J-Y. 2013. Hsp90 regulates nongenetic variation in response to environmental stress. Mol. Cell. 50:82–92. doi: 10.1016/j.molcel.2013.01.026.

117. Hubert DA et al. 2003. Cytosolic HSP90 associates with and modulates the Arabidopsis RPM1 disease resistance protein. EMBO J. 22:5679–5689. doi: 10.1093/emboj/cdg547.

118. Husar S et al. 2011. Overexpression of the UGT73C6 alters brassinosteroid glucoside formation in Arabidopsis thaliana. BMC Plant Biol. 11:51. doi: 10.1186/1471-2229-11-51.

119. Iki T et al. 2010. In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. Mol. Cell. 39:282–291. doi: 10.1016/j.molcel.2010.05.014.

120. Im Kim J et al. 2011. YUCCA6 over-expression demonstrates auxin function in delaying leaf senescence in Arabidopsis thaliana. J. Exp. Bot. 62:3981–3992.

121. Initiative AG. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature. 408:796.

122. Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet. 11:97–108. doi: nrg2689 [pii] 10.1038/nrg2689.

123. Ishiguro S et al. 2002. SHEPARD is the Arabidopsis GRP94 responsible for the formation of functional CLAVATA proteins. EMBO J. 21:898–908.

124. Jack T. 2004. Molecular and genetic mechanisms of floral control. Plant Cell. 16 Suppl:S1–17. doi: 10.1105/tpc.017038.

125. Jaillais Y, Vert G. 2012. Brassinosteroids, gibberellins and light-mediated signalling are the three-way controls of plant sprouting. Nat. Cell Biol. 14:788–790.

126. Jarosz D, Lindquist S. 2010. Hsp90 and environmental stress transform the adaptive value of natural genetic variation. Science (80-.). 330:1820–1824. doi: 330/6012/1820 [pii] 10.1126/science.1195487.

127. Jarosz D, Taipale M, Lindquist SL. 2010. Protein homeostasis and the phenotypic manifestation of genetic diversity: principles and mechanisms. Annu Rev Genet. 44:189–216. doi: 10.1146/annurev.genet.40.110405.090412.

128. Jermy AJ, Willer M, Davis E, Wilkinson BM, Stirling CJ. 2006. The Brl domain in Sec63p is required for assembly of functional endoplasmic reticulum translocons. J. Biol. Chem. 281:7899–906. doi: 10.1074/jbc.M511402200.

129. Jiao Y et al. 2011. Ancestral polyploidy in seed plants and angiosperms. Nature. 473:97–100. doi: 10.1038/nature09916.

130. Jimenez-Gomez JM, Corwin JA, Joseph B, Maloof JN, Kliebenstein DJ. 2011. Genomic analysis of QTLs and genes altering natural variation in stochastic noise Gibson, G, editor. PLoS Genet. 7:e1002295. doi: 10.1371/journal.pgen.1002295 PGENETICS-D-11-00547 [pii].

131. Jin S-H et al. 2013. UGT74D1 is a novel auxin glycosyltransferase from Arabidopsis thaliana. Muday, G, editor. PLoS One. 8:e61705. doi: 10.1371/journal.pone.0061705.

132. Johnson M et al. 2008. NCBI BLAST: a better web interface. Nucleic Acids Res. 36:W5–W9.

133. Kærn M, Elston TC, Blake WJ, Collins JJ. 2005. Stochasticity in gene expression: from theories to phenotypes. Nat. Rev. Genet. 6:451–464.

134. Kampinga HH, Craig EA. 2010. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nat. Rev. Mol. Cell Biol. 11:579–592.

135. Karagöz GE et al. 2014. Hsp90-Tau Complex Reveals Molecular Basis for Specificity in Chaperone Action. Cell. 156:963–974.

136. Karimi M, Inzé D, Depicker A. 2002. GATEWAYTM vectors for *Agrobacterium*-mediated plant transformation. Trends Plant Sci. 7:193–195.

137. Kerner MJ et al. 2005. Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. Cell. 122:209–220. doi: 10.1016/j.cell.2005.05.028.

138. Kim S et al. 2007. Recombination and linkage disequilibrium in Arabidopsis thaliana. Nat. Genet. 39:1151–1155.

139. Kim TS et al. 2011. HSP90 functions in the circadian clock through stabilization of the client F-box protein ZEITLUPE. Proc. Natl. Acad. Sci. U. S. A. 108:16843–16848. doi: 10.1073/pnas.1110406108.

140. Kim T-W, Wang. 2010. Brassinosteroid signal transduction from receptor kinases to transcription factors. Annu. Rev. Plant Biol. 61:681–704. doi: doi:10.1146/annurev.arplant.043008.092057.

141. Kleczkowski K, Schell J, Bandur R. 1995. Phytohormone conjugates: nature and function. CRC. Crit. Rev. Plant Sci. 14:283–298.

142. Koornneef M, Alonso-Blanco C, Peeters AJM, Soppe W. 1998. Genetic control of flowering time in Arabidopsis. Annu. Rev. Plant Biol. 49:345–370.

143. Kramer EM, Jaramillo MA, Di Stilio VS. 2004. Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. Genetics. 166:1011–1023.

144. Kryazhimskiy S, Plotkin JB. 2008. The population genetics of dN/dS. Gojobori, T, editor. PLoS Genet. 4:e1000304. doi: 10.1371/journal.pgen.1000304.

145. Lachowiec J et al. 2013. The protein chaperone HSP90 can facilitate the divergence of gene duplicates. Genetics. 193:1269–77. doi: 10.1534/genetics.112.148098.

146. Lamesch P et al. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 40:D1202–D1210.

147. Larsson AS, Landberg K, Meeks-Wagner DR. 1998. The TERMINAL FLOWER2 (TFL2) Gene Controls the Reproductive Transition and Meristem Identity in Arabidopsis thaliana. Genetics. 149:597–605. http://www.genetics.org/content/149/2/597.full (Accessed December 5, 2013).

148. Leffler EM et al. 2012. Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol. 10:e1001388.

149. Lehner B. 2010. Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast. PLoS One. 5:e9035.

150. Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG. 2006. Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. Nat Genet. 38:896–903. doi: ng1844 [pii] 10.1038/ng1844.

151. Lempe J, Lachowiec J, Sullivan AM, Queitsch C. 2012. Molecular mechanisms of robustness in plants. Curr Opin Plant Biol. doi: S1369-5266(12)00172-0 [pii] 10.1016/j.pbi.2012.12.002.

152. Levy SF, Siegal ML. 2008. Network hubs buffer environmental variation in Saccharomyces cerevisiae Levchenko, A, editor. PLoS Biol. 6:e264. doi: 10.1371/journal.pbio.0060264.

153. Li J, Li Y, Chen S, An L. 2010. Involvement of brassinosteroid signals in the floral-induction network of Arabidopsis. J. Exp. Bot. 61:4221–30. doi: 10.1093/jxb/erq241.

154. Li L et al. 2009. Arabidopsis MYB30 is a direct target of BES1 and cooperates with BES1 to regulate brassinosteroid-induced gene expression. Plant J. 58:275–286. doi: 10.1111/j.1365-313X.2008.03778.x.

155. Li Q-F et al. 2012. An interaction between BZR1 and DELLAs mediates direct signaling crosstalk between brassinosteroids and gibberellins in Arabidopsis. Sci. Signal. 5:ra72.

156. Li W et al. 2008. Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling. PLoS One. 3:e1487.

157. Li Y, Baldauf S, Lim E-K, Bowles DJ. 2001. Phylogenetic analysis of the UDP-glycosyltransferase multigene family of Arabidopsis thaliana. J. Biol. Chem. 276:4338–4343.

158. Li, Ye H, Guo, Yin. 2010. Arabidopsis IWS1 interacts with transcription factor BES1 and is involved in plant steroid hormone brassinosteroid regulated gene expression. Proc. Natl. Acad. Sci. U. S. A. 107:3918–23. doi: 10.1073/pnas.0909198107.

159. Lim WA, Pawson T. 2010. Phosphotyrosine signaling: evolving a new cellular communication system. Cell. 142:661–667.

160. Liu C et al. 2010. Coupled chaperone action in folding and assembly of hexadecameric Rubisco. Nature. 463:197–202.

161. Liu H, Charng Y. 2013. Common and distinct functions of Arabidopsis class A1 and A2 heat shock factors in diverse abiotic stress responses and development. Plant Physiol. 163:276–90. doi: 10.1104/pp.113.221168.

162. Liu S-L, Baute GJ, Adams KL. 2011. Organ and cell type-specific complementary expression patterns and regulatory neofunctionalization between duplicated genes in Arabidopsis thaliana. Genome Biol. Evol. 3:1419–36. doi: 10.1093/gbe/evr114.

163. Logemann E, Birkenbihl RP, Ülker B, Somssich IE. 2006. An improved method for preparing Agrobacterium cells that simplifies the Arabidopsis transformation protocol. Plant Methods. 2:16.

164. De Lucas M, Brady SM. 2013. Gene regulatory networks in the Arabidopsis root. Curr. Opin. Plant Biol. 16:50–55.

165. Lynch M, Conery JS. 2000. The Evolutionary Fate and Consequences of Duplicate Genes. Science (80-.). 290:1151–1155. doi: 10.1126/science.290.5494.1151.

166. Mackay TF. 2001. The genetic architecture of quantitative traits. Annu. Rev. Genet. 35:303–39. doi: 10.1146/annurev.genet.35.102401.090633.

167. Maere S et al. 2005. Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci U S A. 102:5454–5459. doi: 0501102102 [pii] 10.1073/pnas.0501102102.

168. Makumburage GB, Stapleton AE. 2011. Phenotype uniformity in combined-stress environments has a different genetic architecture than in single-stress treatments. Front. Plant Sci. 2:12. doi: 10.3389/fpls.2011.00012.

169. Manning G, Hunter T. 2009. Eukaryotic Kinomes: Genomics and Evolution of Protein Kinases. In: Handbook of Cell Signaling. Dennis, EA & Bradshaw, RA, editors. Academic Press: San Diego pp. 13–17.

170. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. 2002. The protein kinase complement of the human genome. Science. 298:1912–34. doi: 10.1126/science.1075762.

171. Markham JE et al. 2011. Sphingolipids containing very-long-chain fatty acids define a secretory pathway for specific polar plasma membrane protein targeting in Arabidopsis. Plant Cell Online. 23:2362–2378. doi: 10.1105/tpc.110.080473.

172. Martinez A et al. 2006. Epistatic interaction between FCRL3 and NFκB1 genes in Spanish patients with rheumatoid arthritis. Ann. Rheum. Dis. 65:1188–1191. doi: 10.1136/ard.2005.048454.

173. Mas P, Yanovsky MJ. 2009. Time for circadian rhythms: plants get synchronized. Curr. Opin. Plant Biol. 12:574–579.

174. Masel J. 2006. Cryptic genetic variation is enriched for potential adaptations. Genetics. 172:1985–1991.

175. Masel J, Siegal ML. 2009. Robustness: mechanisms and consequences. Trends Genet. 25:395–403.

176. Mathur J et al. 1998. Transcription of the Arabidopsis CPD gene, encoding a steroidogenic cytochrome P450, is negatively controlled by brassinosteroids. Plant J. 14:593–602.

177. McClellan AJ et al. 2007. Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches. Cell. 131:121–135.

178. McClellan AJ, Tam S, Kaganovich D, Frydman J. 2005. Protein quality control: chaperones culling corrupt conformations. Nat Cell Biol. 7:736–741. http://dx.doi.org/10.1038/ncb0805-736.

179. McGuigan K, Nishimura N, Currey M, Hurwit D, Cresko WA. 2011. Cryptic genetic variation and body size evolution in threespine stickleback. Evolution (N. Y). 65:1203–1211.

180. McLellan CA et al. 2007. A Rhizosphere Fungus Enhances Arabidopsis Thermotolerance through Production of an HSP90 Inhibitor. Plant Physiol. 145:174–182. doi: 10.1104/pp.107.101808.

181. Meiklejohn CD, Hartl DL. 2002. A single mode of canalization. Trends Ecol. Evol. 17:468–473. doi: 10.1016/s0169-5347(02)02596-x.

182. Milloz J, Duveau F, Nuez I, Félix M-A. 2008. Intraspecific evolution of the intercellular signaling network underlying a robust developmental system. Genes Dev. 22:3064–3075.

183. Milton CC, Huynh B, Batterham P, Rutherford SL, Hoffmann AA. 2003. Quantitative trait symmetry independent of Hsp90 buffering: distinct modes of genetic canalization and developmental stability. Proc. Natl. Acad. Sci. 100:13396–13401.

184. Milton CC, Ulane CM, Rutherford S. 2006. Control of canalization and evolvability by Hsp90. PLoS One. 1:e75.

185. Morishima Y et al. 2008. CHIP deletion reveals functional redundancy of E3 ligases in promoting degradation of both signaling proteins and expanded glutamine proteins. Hum. Mol. Genet. 17:3942–3952. doi: 10.1093/hmg/ddn296.

186. Motani R, Schmitz L. 2011. Phylogenetic versus functional signals in the evolution of form–function relationships in terrestrial vision. Evolution (N. Y). 65:2245–2257.

187. Mouchel CF, Osmont KS, Hardtke CS. 2006. BRX mediates feedback between brassinosteroid levels and auxin signalling in root growth. Nature. 443:458–461.

188. Murata S, Minami Y, Minami M, Chiba T, Tanaka K. 2001. CHIP is a chaperone-dependent E3 ligase that ubiquitylates unfolded protein. EMBO Rep. 2:1133–1138. doi: 10.1093/embo-reports/kve246.

189. Mustroph A, Sonnewald U, Biemelt S. 2007. Characterisation of the ATP-dependent phosphofructokinase gene family from *Arabidopsis thaliana*. FEBS Lett. 581:2401–2410.

190. Nathan DF, Vos MH, Lindquist SL. 1997. In vivo functions of the Saccharomyces cerevisiae Hsp90 chaperone. Proc. Natl. Acad. Sci. 94:12949–12956. http://www.pnas.org/content/94/24/12949.abstract.

191. Nei M, Roychoudhury AK. 1973. Probability of Fixation of Nonfunctional Genes at Duplicate Loci. Am. Nat. 107:362–372. doi: 10.2307/2459537.

192. Nelson RM, Pettersson ME, Li X, Carlborg Ö. 2013. Variance Heterogeneity in Saccharomyces cerevisiae Expression Data: Trans-Regulation and Epistasis. PLoS One. 8:e79507.

193. Nemhauser JL. 2008. Dawning of a new era: photomorphogenesis as an integrated molecular network. Curr. Opin. Plant Biol. 11:4–8. doi: 10.1016/j.pbi.2007.10.005.

194. Nemhauser JL, Mockler TC, Chory J. 2004. Interdependency of brassinosteroid and auxin signaling in Arabidopsis. PLoS Biol. 2:e258.

195. Nielsen R et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 3:e170.

196. Nusinow DA et al. 2011. The ELF4-ELF3-LUX complex links the circadian clock to diurnal control of hypocotyl growth. Nature. 475:398–402.

197. Oh E, Zhu J-Y, Wang. 2012. Interaction between BZR1 and PIF4 integrates brassinosteroid and environmental responses. Nat. Cell Biol. 14:802–809.

198. Ordas B, Malvar RA, Hill WG. 2008. Genetic variation and quantitative trait loci associated with developmental stability and the environmental correlation between traits in maize. Genet Res. 90:385–395. doi: 10.1017/S0016672308009762 S0016672308009762 [pii].

199. Ossowski S et al. 2008. Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res. 18:2024–2033.

200. Paaby AB, Rockman M V. 2014. Cryptic genetic variation: evolution's hidden substrate. Nat. Rev. Genet.

201. Pagel M. 1994. Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters. Proc. Biol. Sci. 255:37–45.

202. Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature. 401:877–884. http://dx.doi.org/10.1038/44766.

203. Pagel M, Meade A, Barker D. 2004. Bayesian estimation of ancestral character states on phylogenies. Syst. Biol. 53:673–84. doi: 10.1080/10635150490522232.

204. Pál C, Papp BB, Hurst LD. 2001. Highly Expressed Genes in Yeast Evolve Slowly. Genetics. 158:927–931. http://www.genetics.org/content/158/2/927.short (Accessed December 1, 2013).

205. Papp B, Pal C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene families in yeast. Nature. 424:194–197. doi: 10.1038/nature01771 nature01771 [pii].

206. Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 20:289–290.

207. Park K, Park J, Kim J, Kwak I-S. 2010. Biological and molecular responses of *Chironomus riparius* (Diptera, Chironomidae) to herbicide 2, 4-D (2, 4-dichlorophenoxyacetic acid). Comp. Biochem. Physiol. Part C Toxicol. Pharmacol. 151:439–446.

208. Parkinson SE, Gross SM, Hollick JB. 2007. Maize sex determination and abaxial leaf fates are canalized by a factor that maintains repressed epigenetic states. Dev. Biol. 308:462–473.

209. Paterson AH, Saranga Y, Menz M, Jiang C-X, Wright R. 2003. QTL analysis of genotype× environment interactions affecting cotton fiber quality. Theor. Appl. Genet. 106:384–396.

210. Pena MI, Davlieva M, Bennett MR, Olson JS, Shamoo Y. 2010. Evolutionary fates within a microbial population highlight an essential role for protein folding during natural selection. Mol Syst Biol. 6:387. doi: msb201043 [pii] 10.1038/msb.2010.43.

211. Perry L. 2002. Starch Granule Size and the Domestication of Manioc (Manihot Esculenta) and Sweet Potato (Ipomoea Batatas)1. Econ. Bot. 56:335–349. doi: 10.1663/0013-0001(2002)056[0335:SGSATD]2.0.CO;2.

212. Pesch M, Hülskamp M. 2004. Creating a two-dimensional pattern *de novo* during *Arabidopsis* trichome and root hair initiation. Curr. Opin. Genet. Dev. 14:422–427.

213. Pettersson ME, Nelson RM, Carlborg O. 2012. Selection on variance-controlling genes: adaptability or stability. Evolution (N. Y). 66:3945–3949. doi: 10.1111/j.1558-5646.2012.01753.x.

214. Picard D et al. 1990. Reduced levels of hsp90 compromise steroid receptor action in vivo. Nature. 348:166–168. doi: 10.1038/348166a0.

215. Poppenberger B et al. 2005. The UGT73C5 of Arabidopsis thaliana glucosylates brassinosteroids. Proc. Natl. Acad. Sci. U. S. A. 102:15253–8. doi: 10.1073/pnas.0504279102.

216. Pratt CW, Cornely K. 2012. Essential biochemistry. Wiley Global Education.

217. Pratt WB. 1990. Interaction of hsp90 with steroid receptors: organizing some diverse observations and presenting the newest concepts. Mol. Cell. Endocrinol. 74:C69–C76.

218. Pratt WB, Morishima Y, Peng H-M, Osawa Y. 2010. Proposal for a role of the Hsp90/Hsp70-based chaperone machinery in making triage decisions when proteins undergo oxidative and toxic damage. Exp. Biol. Med. 235:278–289.

219. Press MO et al. 2013. Genome-scale Co-evolutionary Inference Identifies Functions and Clients of Bacterial Hsp90. PLoS Genet. 9:e1003631. doi: 10.1371/journal.pgen.1003631.

220. Priest DM, Jackson RG, Ashford DA, Abrams SR, Bowles DJ. 2005. The use of abscisic acid analogues to analyse the substrate selectivity of UGT71B6, a UDP-glycosyltransferase of Arabidopsis thaliana. http://www.sciencedirect.com/science/article/pii/S0014579305008537 (Accessed December 3, 2013).

221. Pujato M, MacCarthy T, Fiser A, Bergman A. 2013. The underlying molecular and network level mechanisms in the evolution of robustness in gene regulatory networks. PLoS Comput. Biol. 9:e1002865.

222. Purcell S et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.

223. Pyun J-A, Kim S, Cha DH, Kwack K. 2013. Epistasis between polymorphisms in TSHB and ADAMTS16 is associated with premature ovarian failure. Menopause (New York, NY).

224. Queitsch C, Carlson KD, Girirajan S. 2012. Lessons from model organisms: phenotypic robustness and missing heritability in complex disease. PLoS Genet. 8:e1003041. doi: 10.1371/journal.pgen.1003041 PGENETICS-D-12-00749 [pii].

225. Queitsch C, Sangster T, Lindquist S. 2002. Hsp90 as a capacitor of phenotypic variation. Nature. 417:618–624. doi: 10.1038/nature749 nature749 [pii].

226. Rando OJ, Verstrepen KJ. 2007. Timescales of genetic and epigenetic inheritance. Cell. 128:655–668. doi: 10.1016/j.cell.2007.01.023.

227. Rinott R, Jaimovich A, Friedman N. 2011. Exploring transcription regulation through cell-to-cell variability. Proc. Natl. Acad. Sci. U. S. A. 108:6329–34. doi: 10.1073/pnas.1013148108.

228. Robert-Seilaniantz A, Grant M, Jones JD. 2011. Hormone crosstalk in plant disease and defense: more than just jasmonate-salicylate antagonism. Annu. Rev. Phytopathol. 49:317–343. doi: 10.1146/annurev-phyto-073009-114447.

229. Rogers S, Wells R, Rechsteiner M. 1986. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science (80-.). 234:364–368.

230. Röhl A, Rohrberg J, Buchner J. 2013. The chaperone Hsp90: changing partners for demanding clients. Trends Biochem. Sci. 38:253–262.

231. Rohner N et al. 2013. Cryptic Variation in Morphological Evolution: HSP90 as a Capacitor for Loss of Eyes in Cavefish. Science (80-.). 342:1372–1375.

232. Rönnegård L, Shen X, Alam M. 2010. hglm: A Package for Fitting Hierarchical Generalized Linear Models. R J. 2.

233. Rozhon W, Mayerhofer J, Petutschnig E, Fujioka S, Jonak C. 2010. ASKθ, a group-III Arabidopsis GSK3, functions in the brassinosteroid signalling pathway. Plant J. 62:215–223.

234. Rutherford SL, Lindquist SL. 1998. Hsp90 as a capacitor for morphological evolution. Nature. 396:336–342. doi: 10.1038/24550.

235. Salverda MLM et al. 2011. Initial mutations direct alternative pathways of protein evolution. PLoS Genet. 7:e1001321. http://dx.doi.org/10.1371/journal.pgen.1001321.

236. Samakovli D, Thanou A, Valmas C, Hatzopoulos P. 2007. Hsp90 canalizes developmental perturbation. J. Exp. Bot. 58:3513–3524.

237. SanCristobal-Gaudy M, Bodin L, Elsen J-M, Chevalet C. 2001. Genetic components of litter size variability in sheep. Genet. Sel. Evol. 33:249–272.

238. Sangster T et al. 2008. HSP90-buffered genetic variation is common in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U. S. A. 105:2969–2974. doi: 10.1073/pnas.0712210105.

239. Sangster TA et al. 2008. HSP90 affects the expression of genetic variation and developmental stability in quantitative traits. Proc. Natl. Acad. Sci. U. S. A. 105:2963–2968. doi: 10.1073/pnas.0712200105.

240. Sangster TA et al. 2007. Phenotypic diversity and altered environmental plasticity in Arabidopsis thaliana with reduced Hsp90 levels. PLoS One. 2:e648. doi: 10.1371/journal.pone.0000648.

241. Sangster TA. 2007. Quantitative genetic analysis of HSP90-buffered variation. THE UNIVERSITY OF CHICAGO.

242. Sangster TA, Lindquist SL, Queitsch C. 2004. Under cover: causes, effects and implications of Hsp90-mediated genetic capacitance. Bioessays. 26:348–362. doi: 10.1002/bies.20020.

243. Sangster TA, Queitsch C. 2005. The HSP90 chaperone complex, an emerging force in plant development and phenotypic plasticity. Curr Opin Plant Biol. 8:86–92. doi: S1369-5266(04)00164-5 [pii] 10.1016/j.pbi.2004.11.012.

244. Santelia D, Zeeman SC. 2011. Progress in Arabidopsis starch research and potential biotechnological applications. Curr. Opin. Biotechnol. 22:271–280.

245. Sapkota Y et al. 2013. Assessing SNP-SNP interactions among DNA repair, modification and metabolism related pathway genes in breast cancer susceptibility. PLoS One. 8:e64896.

246. Scheible W-R, Pauly M. 2004. Glycosyltransferases and cell wall biosynthesis: novel players and insights. Curr. Opin. Plant Biol. 7:285–295.

247. Scheres B, Benfey P, Dolan L. 2002. Root development. Arabidopsis Book. 1:e0101. doi: 10.1199/tab.0101.

248. Schleiff E, Becker T. 2011. Common ground for protein translocation: access control for mitochondria and chloroplasts. Nat. Rev. Mol. Cell Biol. 12:48–59. doi: 10.1038/nrm3027.

249. Schmid M et al. 2005. A gene expression map of Arabidopsis thaliana development. Nat. Genet. 37:501–506.

250. Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 9:671–675.

251. Sekimoto T et al. 2010. The molecular chaperone Hsp90 regulates accumulation of DNA polymerase η at replication stalling sites in UV-irradiated cells. Mol. Cell. 37:79–89.

252. Sgrò CM, Wegener B, Hoffmann AA. 2010. A naturally occurring variant of Hsp90 that is associated with decanalization. Proc. R. Soc. B Biol. Sci. 277:2049–2057.

253. Shen X et al. 2012. Inheritance beyond plain heritability: variance-controlling genes in Arabidopsis thaliana. Barsh, GS, editor. PLoS Genet. 8:e1002839. doi: 10.1371/journal.pgen.1002839.

254. Shen X, Alam M, Fikse F, Rönnegård L. 2013. A novel generalized ridge regression method for quantitative genetics. Genetics. 193:1255–1268.

255. Shigeta T et al. 2013. Molecular evidence of the involvement of heat shock protein 90 in brassinosteroid signaling in Arabidopsis T87 cultured cells. Plant Cell Rep. 33:1–12. doi: 10.1007/s00299-013-1550-y.

256. Shpak ED, Berthiaume CT, Hill EJ, Torii KU. 2004. Synergistic interaction of three ERECTA-family receptor-like kinases controls Arabidopsis organ growth and flower development by promoting cell proliferation. Development. 131:1491–1501.

257. Shpak ED, McAbee JM, Pillitteri LJ, Torii KU. 2005. Stomatal patterning and differentiation by synergistic interactions of receptor kinases. Science (80-.). 309:290–293.

258. Sieber P, Wellmer F, Gheyselinck J, Riechmann JL, Meyerowitz EM. 2007. Redundancy and specialization among plant microRNAs: role of the MIR164 family in developmental robustness. Development. 134:1051–1060.

259. Siegal ML, Bergman A. 2002. Waddington's canalization revisited: developmental stability and evolution. Proc. Natl. Acad. Sci. U. S. A. 99:10528–10532. doi: 10.1073/pnas.102303999.

260. Simillion C, Vandepoele K, Van Montagu MCE, Zabeau M, Van de Peer Y. 2002. The hidden duplication past of Arabidopsis thaliana. Proc. Natl. Acad. Sci. 99:13627–13632.

261. De Smet R et al. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc. Natl. Acad. Sci. U. S. A. 110:2898–903. doi: 10.1073/pnas.1300127110.

262. Sohn EJ et al. 2007. The shoot meristem identity gene TFL1 is involved in flower development and trafficking to the protein storage vacuole. Proc. Natl. Acad. Sci. U. S. A. 104:18801–6. doi: 10.1073/pnas.0708236104.

263. Sollars V et al. 2003. Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. Nat. Genet. 33:70–74. doi: 10.1038/ng1067.

264. Specchia V et al. 2010. Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. Nature. 463:662–665.

265. Stensløkken K-O, Ellefsen S, Larsen HK, Vaage J, Nilsson GE. 2010. Expression of heat shock proteins in anoxic crucian carp (Carassius carassius): support for cold as a preparatory cue for anoxia. Am. J. Physiol. Integr. Comp. Physiol. 298:R1499–R1508.

266. Stepanova AN, Yun J, Likhacheva A V, Alonso JM. 2007. Multilevel interactions between ethylene and auxin in Arabidopsis roots. Plant Cell Online. 19:2169–2185.

267. Stewart JL, Maloof JN, Nemhauser JL. 2011. PIF genes mediate the effect of sucrose on seedling growth dynamics. PLoS One. 6:e19894.

268. Sun et al. 2010. Integration of brassinosteroid signal transduction with the transcription network for plant growth regulation in Arabidopsis. Dev. Cell. 19:765–777. doi: 10.1016/j.devcel.2010.10.010.

269. Swanson WJ, Yang Z, Wolfner MF, Aquadro CF. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc. Natl. Acad. Sci. 98:2509–2514.

270. Swarbreck D et al. 2008. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res. 36:D1009–D1014. doi: 10.1093/nar/gkm965.

271. Swarup R et al. 2007. Ethylene upregulates auxin biosynthesis in Arabidopsis seedlings to enhance inhibition of root cell elongation. Plant Cell Online. 19:2186–2196.

272. Taipale M et al. 2012. Quantitative analysis of HSP90-client interactions reveals principles of substrate recognition. Cell. 150:987–1001. doi: 10.1016/j.cell.2012.06.047 S0092-8674(12)00937-3 [pii].

273. Taipale M, Jarosz D, Lindquist S. 2010. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. Nat Rev Mol Cell Biol. 11:515–528. doi: nrm2918 [pii] 10.1038/nrm2918.

274. Takahashi A, Casais C, Ichimura K, Shirasu K. 2003. HSP90 interacts with RAR1 and SGT1 and is essential for RPS2-mediated disease resistance in Arabidopsis. Proc Natl Acad Sci U S A. 100:11777–11782. doi: 10.1073/pnas.2033934100 2033934100 [pii].

275. Tang W et al. 2011. PP2A activates brassinosteroid-responsive gene expression and plant growth by dephosphorylating BZR1. Nat. Cell Biol. 13:124–131.

276. Tennessen JA et al. 2012. Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. Science (80-.). 337:64–69. doi: 10.1126/science.1219240.

277. Tokuriki N, Tawfik DS. 2009. Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature. 459:668–673. doi: nature08009 [pii] 10.1038/nature08009.

278. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. 2009. A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. 10:252–263. doi: http://www.nature.com/nrg/journal/v10/n4/suppinfo/nrg2538_S1.html.

279. Vaucheret H et al. 1998. Transgene-induced gene silencing in plants. Plant J. 16:651–659.

280. Vaughan CK, Neckers L, Piper PW. 2010. Understanding of the Hsp90 molecular chaperone reaches new heights. Nat. Struct. Mol. Biol. 17:1400–1404. http://dx.doi.org/10.1038/nsmb1210-1400.

281. Vert G, Chory J. 2006. Downstream nuclear events in brassinosteroid signalling. Nature. 441:96–100. doi: 10.1038/nature04681.

282. Waddington CH. 1942. Canalization of development and the inheritance of acquired characters. Nature. 150:563–565.

283. Wagner A. 2005. Distributed robustness versus redundancy as causes of mutational robustness. Bioessays. 27:176–88. doi: 10.1002/bies.20170.

284. Wagner A. 1996. Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. Biol. Cybern. 74:557–567.

285. Wagner A. 2000. Robustness against mutations in genetic networks of yeast. Nat. Genet. 24:355–361.

286. Walcher CL, Nemhauser JL. 2012. Bipartite promoter element required for auxin response. Plant Physiol. 158:273–282.

287. Wall DPP et al. 2005. Functional genomic analysis of the rates of protein evolution. Proc Natl Acad Sci U S A. 102:5483–5488. doi: 0501761102 [pii] 10.1073/pnas.0501761102.

288. Wang et al. 2002. Nuclear-localized BZR1 mediates brassinosteroid-induced growth and feedback suppression of brassinosteroid biosynthesis. Dev. Cell. 2:505–513. http://www.ncbi.nlm.nih.gov/pubmed/11970900 (Accessed April 24, 2014).

289. Wang Y et al. 2013. Strigolactone/MAX2-Induced Degradation of Brassinosteroid Transcriptional Effector BES1 Regulates Shoot Branching. Dev. Cell. 27:681–688.

290. Wang ZY, Seto H, Fujioka S, Yoshida S, Chory J. 2001. BRI1 is a critical component of a plasma-membrane receptor for plant steroids. Nature. 410:380–383. doi: 10.1038/35066597 35066597 [pii].

291. Wapinski I, Pfeffer A, Friedman N, Regev A. 2007. Natural history and evolutionary principles of gene duplication in fungi. Nature. 449:54–61. doi: nature06107 [pii] 10.1038/nature06107.

292. Wegele H, Muller L, Buchner J. 2004. Hsp70 and Hsp90--a relay team for protein folding. Rev Physiol Biochem Pharmacol. 151:1–44. doi: 10.1007/s10254-003-0021-1.

293. Whitacre JM. 2012. Biological robustness: paradigms, mechanisms, and systems principles. Front. Genet. 3:67.

294. Whitesell L, Mimnaugh EG, De Costa B, Myers CE, Neckers LM. 1994. Inhibition of heat shock protein HSP90-pp60v-src heteroprotein complex formation by benzoquinone ansamycins: essential role for stress proteins in oncogenic transformation. Proc Natl Acad Sci U S A. 91:8324–8328. http://www.ncbi.nlm.nih.gov/pubmed/8078881.

295. Williams TA, Fares MA. 2010. The effect of chaperonin buffering on protein evolution. Genome Biol. Evol. 2:609–619. doi: 10.1093/gbe/evq045.

296. Winter D et al. 2007. An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. PLoS One. 2:e718.

297. Workman P, Burrows F, Neckers LEN, Rosen N. 2007. Drugging the cancer chaperone HSP90. Ann. N. Y. Acad. Sci. 1113:202–216.

298. Wright SI, Yau CB, Looseley M, Meyers BC. 2004. Effects of gene expression on molecular evolution in Arabidopsis thaliana and Arabidopsis lyrata. Mol Biol Evol. 21:1719–1726. doi: 10.1093/molbev/msh191 msh191 [pii].

299. Wu CI, Li WH. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. 82:1741–1745. doi: 10.1073/pnas.82.6.1741.

300. Xu Z, Horwich AL, Sigler PB. 1997. The crystal structure of the asymmetric GroEL–GroES–(ADP) 7 chaperonin complex. Nature. 388:741–750.

301. Yang J et al. 2012. FTO genotype is associated with phenotypic variability of body mass index. Nature. 490:267–72. doi: 10.1038/nature11401.

302. Yeyati PLL, Bancewicz RMM, Maule J, Van Heyningen V. 2007. Hsp90 selectively modulates phenotype in vertebrate development. PLoS Genet. 3:e43. doi: 06-PLGE-RA-0249R4 [pii] 10.1371/journal.pgen.0030043.

303. Yin et al. 2005. A new class of transcription factors mediates brassinosteroid-regulated gene expression in Arabidopsis. Cell. 120:249–259. doi: 10.1016/j.cell.2004.11.044.

304. Yin et al. 2002. BES1 accumulates in the nucleus in response to brassinosteroids to regulate gene expression and promote stem elongation. Cell. 109:181–191. doi: 10.1016/S0092-8674(02)00721-3.

305. Young JC, Hoogenraad NJ, Hartl FU. 2003. Molecular Chaperones Hsp90 and Hsp70 Deliver Preproteins to the Mitochondrial Import Receptor Tom70. Cell. 112:41–50. doi: 10.1016/S0092-8674(02)01250-3.

306. Yu X et al. 2011. A brassinosteroid transcriptional network revealed by genome-wide identification of BESI target genes in Arabidopsis thaliana. Plant J. 65:634–646. doi: 10.1111/j.1365-313X.2010.04449.x.

307. Yu X, Li LL, Guo M, Chory J, Yin. 2008. Modulation of brassinosteroid-regulated gene expression by Jumonji domain-containing proteins ELF6 and REF6 in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 105:7618–23. doi: 10.1073/pnas.0802254105.

308. Zhang J. 2003. Evolution by gene duplication: an update. Trends Ecol. Evol. 18:292–298. doi: 10.1016/S0169-5347(03)00033-8.

309. Zhao J et al. 2002. Two putative BIN2 substrates are nuclear components of brassinosteroid signaling. Plant Physiol. 130:1221–1229.

310. Zhao R et al. 2005. Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. Cell. 120:715–727. doi: S0092-8674(04)01249-8 [pii] 10.1016/j.cell.2004.12.024.

311. Zhiponova MK et al. 2014. Helix–loop–helix/basic helix–loop–helix transcription factor network represses cell elongation in Arabidopsis through an apparent incoherent feed-forward loop. Proc. Natl. Acad. Sci. 111:2824–2829.

312. Zou J, Guo Y, Guettouche T, Smith DF, Voellmy R. 1998. Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1. Cell. 94:471–480. doi: 10.1016/s0092-8674(00)81588-3.

313. Zulawski M, Braginets R, Schulze WX. 2013. PhosPhAt goes kinases—searchable protein kinase target information in the plant phosphorylation site database PhosPhAt. Nucleic Acids Res. 41:D1176–D1184.