Transposable Element-Mediated Balancing Selection at *Hsp90* Underlies Embryo Developmental Variation

Bing Chen,*[,1] Bo Zhang,‡[,1] Lingling Xu,[1] Qing Li,[1] Feng Jiang,[2] Pengcheng Yang,[2] Yanan Xu,[1] and Le Kang*[,1,2]

[1]State Key Laboratory of Integrated Management of Pest Rodents and Insects, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

[2]Beijing Institute of Life Science, Chinese Academy of Sciences, Beijing, China

‡Present address: Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China

***Corresponding authors**: E-mails: lkang@ioz.ac.cn; chenbing@ioz.ac.cn.

Associate editor: Yuseob Kim

Abstract

Understanding the roles of transposable elements (TEs) in the evolution of genome and adaptation is a long-sought goal. Here, we present a new model of TE co-option, in which a TE is harnessed by an essential gene and confers local adaptation through heterozygote advantage. We characterized a human *Alu*-like TE family, the *Lm1* elements, in the genome of the migratory locust *Locusta migratoria* that harbors 0.7 million copies of the elements. Scanning *Lm1* insertions in the natural locust populations revealed the widespread high polymorphism of *Lm1*. An *Lm1* was recruited into the coding region of Heat-shock protein 90 (*Hsp90*), an important molecular chaperone for diverse signal transduction and developmental pathways. Only heterozygotes of the allele are present in natural populations. Allele frequency increases with decreased latitudes in east coastal China, even increasing up to 76% in southern populations. Regions flanking the *Lm1* insertion display clear signatures of a selective sweep linked to *Lm1*. The *Lm1*-mediated *Hsp90* mutation is consequential for the embryonic development of locust. Heterozygous embryos develop faster than the wild type, particularly when cued by long-day parental photoperiod. The heterozygotes also present a reduced within-population variation in embryonic development, i.e., high developmental synchrony of embryos. The naturally occurring *Hsp90* mutation could facilitate multivoltinism and developmental synchronization of the locust in southern tropical region. These results revealed a genetic mechanism behind microevolutionary changes in which balancing selection may have acted to maintain the heterozygote advantage through TE co-option in essential genes.

Key words: **balancing selection, Hsp90, transposable element, selective sweep, embryonic development, heterozygote advantage.**

Article

Introduction

Eukaryotic genomes contain thousands to millions of copies of transposable elements (TEs) and other repetitive elements. For example, the majority of human genome (>66%) is composed by repetitive and repeat-derived elements (de Koning et al. 2011). The *Alu* repeats contain over one million copies that comprise 10% of the genome (Batzer and Deininger 2002). Genome size evolution is associated with TE expansion in the host (Ågren and Wright 2011; Sessegolo et al. 2016). As an example, the large genome (~6.5 gigabases) of the migratory locust (*Locusta migratoria*) is saturated with TEs that constitute ~ 60% of the genome (Wang et al. 2014). The persistence of huge amount of TEs in the genome as well as the effect of TEs on genome evolution and function are of great interest to scientists in the past decades.

Most of TEs became "fossils", lost their activity in evolution, and thus were considered as unnecessary "junk" DNA (Charlesworth and Langley 1989; Biemont and Vieira 2006). However, evidences are found stating that sometimes these elements exert strong effects on genome evolution, including increase in recombination and unequal crossover (Britten 2010), proliferation in copy numbers (Schmidt et al. 2010), and rewiring of regulatory networks (Feschotte 2008; Ellison and Bachtrog 2013; Guio et al. 2014). Genomic landscape of individuals in natural populations could be reshaped by insertion of TEs, which was revealed in few species, such as in humans (Batzer and Deininger 2002; Ewing and Kazazian 2010) and *Drosophila melanogaster* (Gonzalez et al. 2010; Petrov et al. 2011; Begun et al. 2012).

TEs can act as a potent agent mediating the genetic variation and driving phenotypic evolution (Ullastres et al. 2015). TE may donate regulatory sequences (Gonzalez et al. 2009; Guio et al. 2014; Ding et al. 2016; van't Hof et al. 2016), and change gene structures of nearby genes (Aminetzach et al. 2005; Feschotte 2008; Rebollo et al. 2012). For example, *P* elements are widespread in proximal promoter of different Heat-shock protein (Hsp) genes in wild *D. melanogaster* flies (Walser et al. 2006). The insertion of *P* elements alters *Hsp* expression and is consequential for thermal stress tolerance and development (Chen et al. 2007; Chen and Wagner 2012). Insertion of a *Doc* TE (*Doc1420*) interrupts the gene

CHKov1, alters the gene transcripts, and is associated with increased pesticide resistance (Aminetzach et al. 2005). TE insertion can also alter regulatory structures of nearby genes (Mateo et al. 2014; Merenciano et al. 2016). In addition, TE insertions can cause a selective sweep with reduced genetic variability among the alleles linked to a TE (Aminetzach et al. 2005; Gonzalez et al. 2008; Gonzalez et al. 2010). However, the influence of genetic variability on phenotypic expression remains less unexplored.

Lm1 element is a human *Alu*-like Short Interspersed Element (SINE), with the presence of millions of copies in the migratory locust genome (Bradfield et al. 1985). *Lm1*-like sequences are also present in other orthopteran species (Wu et al. 2001; Aguirre et al. 2010; Berthier et al. 2011). Similar SINE sequences could occur widely among eukaryotes (Bradfield et al. 1985; Gilbert and Labuda 1999). *Lm1* elements are associated with transcription (Zhang and Wyatt 1996; Patel et al. 2001) and DNA methylation in the migratory locusts (Robinson et al. 2016). However, the roles *Lm1* elements play in genome function and evolution of the locust remain elusive.

We thus conducted a genome-wide investigation of *Lm1* families and their insertion polymorphism in natural populations of the locust. We focused on functional analysis of *Lm1* in *Hsp* genes because Hsps are important molecular chaperones and play an essential role in development and response to environmental stress (Feder and Hofmann 1999; Kang et al. 2009). Hsps consist of several families based on their molecular weight, e.g., Hsp90, Hsp70, and small Hsps (Feder and Hofmann 1999; Huang et al. 2008). Among these families, Hsp90 is predominantly present in the cytosol for both normal and stressful conditions. The functional cytoplasmic Hsp90 is required for viability of eukaryotes under all tested conditions (Borkovich et al. 1989). Different from other Hsps, Hsp90 is associated with specific proteins involved in diverse signal transduction and developmental pathways (Rutherford et al. 2007). For example, Hsp90 plays an essential role in the regulation of embryonic development of insects through binding with its substrates and cochaperones (Bishop et al. 2002; Song et al. 2007; Bradley et al. 2012; King and MacRae 2015). More importantly, Hsp90 can buffer genetic and epigenetic variation whose expression leads to altered phenotypes (Rutherford and Lindquist 1998; Sollars et al. 2003; Chen and Wagner 2012). However, the contribution of Hsp90 to phenotypic evolution in nature is seldom documented (Chen and Wagner 2012; Rohner et al. 2013). Some instances of TE-mediated *Hsp* mutations were recorded in natural populations of *D. melanogaster* (Walser et al. 2006; Chen and Wagner 2012). Nevertheless, these mutations were repressed at extremely low frequency or not independently occurring (Walser et al. 2006; Sgrò et al. 2008; Chen and Wagner 2012), thus devoid of adaptive significance.

In the present study, we discovered the first naturally occurring mutation of *Hsp90* that has swept to high frequency in wide geographic clines. This *Hsp90* variant is an *Lm1* insertional mutation which presents only as heterozygote. The insertion site represents a clear signature of selective sweep linked to the *Lm1*. Interestingly, the *Hsp90* mutation is associated with a photoperiod-dependent developmental change of embryos. This allele could mediate the latitudinal variation in life history traits (e.g., voltinism) of the locusts. Our study provides a model of TE co-option in which TE is harnessed by *Hsp90* gene to adapt to local environments through balancing selection. The findings have implications for genetic mechanisms underlying the microevolution of phenotypic novelty under changing environments.

Results

Lm1 Elements in the Locust Genome: Abundance, Diversity, and Insertion Polymorphism

Canonical *Lm1* elements are a 200-base pair (bp) SINE sequence. Each element contains two 11-bp boxes, box A (TGGCCGAGGTG) and box B (GGTTCGAATCC), followed by a core SINE sequence (fig. 1*A*). The two boxes are highly conserved fragments in eukaryote SINE families and homologous to the split RNA polymerase III (Pol III) promoter of tRNA genes (Gilbert and Labuda 1999).

Lm1 elements are highly abundant in the locust genome. Homolog search identified approximately 715,000 *Lm1*-like sequences in the locust reference genome (BLAST E value: $<$ 1E-20). Thus, the *Lm1* elements comprise an estimated 2.2% of the locust genome.

We examined the gene localization of these *Lm1* elements in the whole locust genome. 91% of *Lm1* are located in the intergenic region; 6% of *Lm1* are in the intronic region; and few *Lm1* ($<$ 0.1%) are in the exonic region (fig. 1*B*).

After filtering the redundant sequences based on their sequence similarity, we obtained 2,133 *Lm1* clusters with at least 10 similar sequences, and 343 *Lm1* clusters with more than 100 similar sequences (see supplementary table S1, Supplementary Material online). The largest cluster consists of 65,917 sequences. We extracted the representative sequences for the first 1,000 largest *Lm1* clusters and used the sequences for alignment and evolutionary analysis (fig. 1*C*). The estimated average evolutionary divergence is 0.156. These *Lm1*-like sequences can be classified into at least seven big subfamilies termed as *Lm1A* to *Lm1G*.

We further investigated the natural insertion polymorphism of *Lm1* elements in six locust populations. The populations are distributed between latitude 18°N and 41°N along the east coast of China (see supplementary table S2, Supplementary Material online). A total of 24 loci located in gene region (intronic or exonic region) were selected for inspecting the insertion polymorphism of *Lm1* belonging to *Lm1A*. Twenty loci exhibit *Lm1* insertion polymorphism ($>$ 5% in maximum insertion frequency) (fig. 1*D* and see supplementary table S3, Supplementary Material online). Half of the *Lm1* insertions remain at low frequency ($<$ 10%), although some *Lm1* insertions range in high allele frequency of up to 57% (i.e., the locus 4 in LN41 population).

Lm1-Mediated *Hsp90* Mutation: Widespread Occurrence, Transcript Alteration, and Heterozygote Predominance

To investigate the possible adaptive effects of *Lm1* insertion, we screened the locust populations for *Lm1* insertion

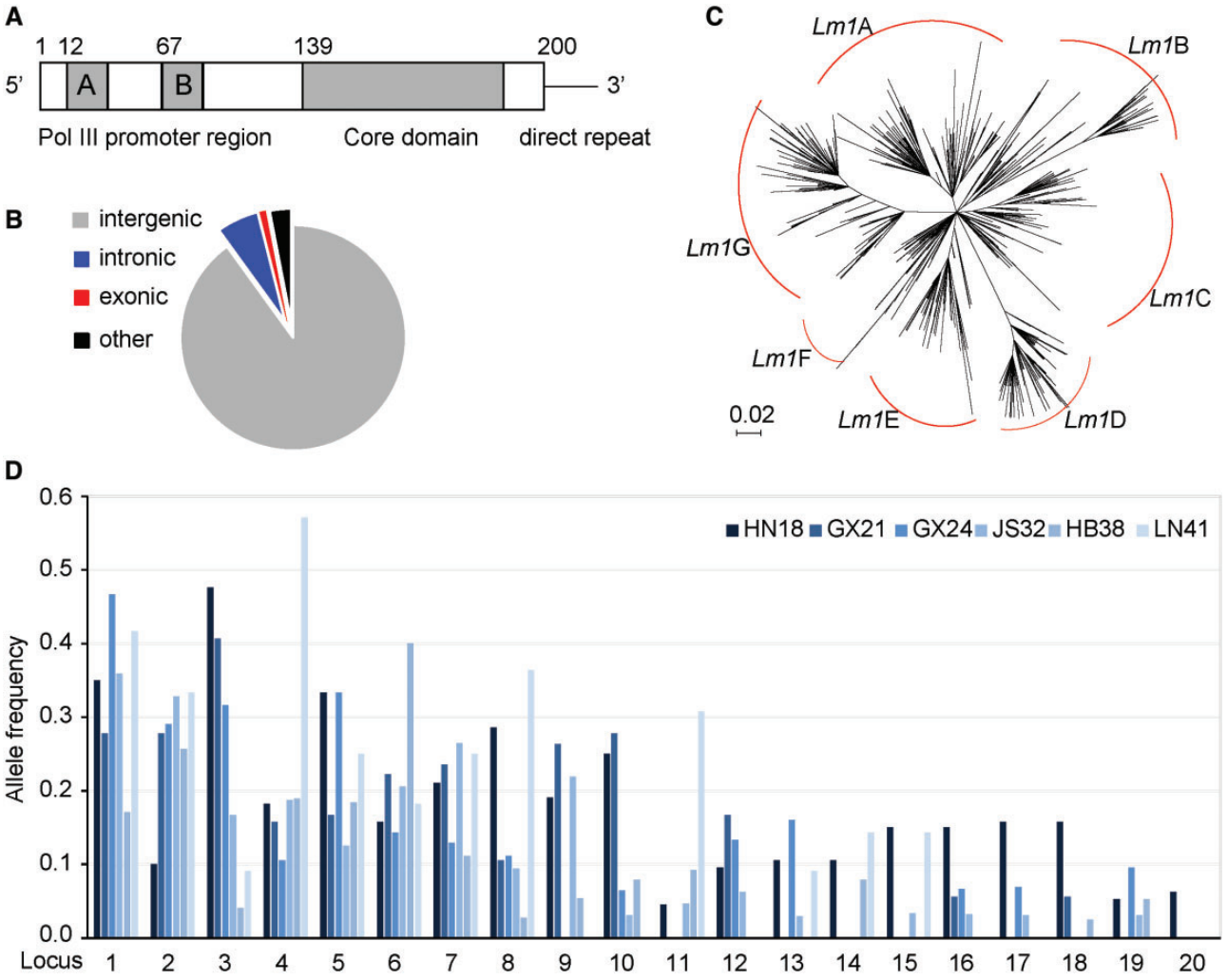

Fig. 1. Localization, diversity, and insertion polymorphism of *Lm1* elements in the locust genome. (*A*) Schematic structure of canonical *Lm1* element. Boxes A and B represent the 11-bp elements of the split RNA Pol III promoter. *Lm1* insertion causes a direct repeat of the flanking sequence. (*B*) Genic localization of all *Lm1* sequences in the reference genome of the locust. (*C*) Evolutionary tree of *Lm1* sequences. The representative sequences for the first 1000 largest *Lm1* clades were used in the phylogenetic analysis. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model. (*D*) *Lm1* insertion polymorphism in six locust populations in 20 loci. The natural populations are from localities ranging from latitude 18°N to 41°N. See the sampling sites and the polymorphic loci in supplementary table S2–S3, Supplementary Material online.

mutations in *Hsp* genes. Four *Hsp* genes from different Hsp families, namely, *Hsp90*, *Hsp70*, *Hsp40*, and *Hsp20.5*, were subject to the inspection. *Lm1* insertion polymorphism was identified in *Hsp90*, *Hsp70* and *Hsp20.5*. The *Lm1* insertions in *Hsp70* and *Hsp20.5* are maintained at low frequency ($<=$10%) (see supplementary table S4, Supplementary Material online). In contrast, the *Lm1* insertion in the single-copy gene *Hsp90* is present in all the natural populations between latitude 18°N and 41°N and maintained at high frequency (more than 71%) in southern populations (fig. 2*A* and see supplementary table S1, Supplementary Material online). *Lm1* sequences in *Hsp90* from these populations represent almost the same sequence belonging to subfamily *Lm1*A (only one SNP), implying that their origin is of the same insertion event.

The *Lm1* insertion in *Hsp90* disrupts the coding region of *Hsp90* and consequently leads to amino acid deletion. The *Lm1* element is inserted into the third exon of *Hsp90* at a position enriched with A_nG_n repeats. The insertion leads to a direct repeat of 13-bp fragment (i.e., GAGGAAGAGGAAG) that flanks the element. We analyzed the amplified cDNA transcripts from at least 20 different *Hsp90* mutant and wild-type mRNAs. The result indicates that the *Lm1* element in the mutant is spliced out during transcription, and not incorporated into the mature mRNA of *Hsp90*. Instead, the mutant transcript has a deletion of 6-bp nucleotides (GAAGAG) at the insertion site. These nucleotides encode two amino acids of glutamate (abbreviated as E) at 227 and 228 of wild-type *Hsp90* protein sequence (fig. 2*B*). This mutation site is located in the charged segment that links the N-terminal domain and middle segment of the protein. The linker region is not conserved among species relative to other regions of Hsp90 (see supplementary fig. S1, Supplementary Material online).

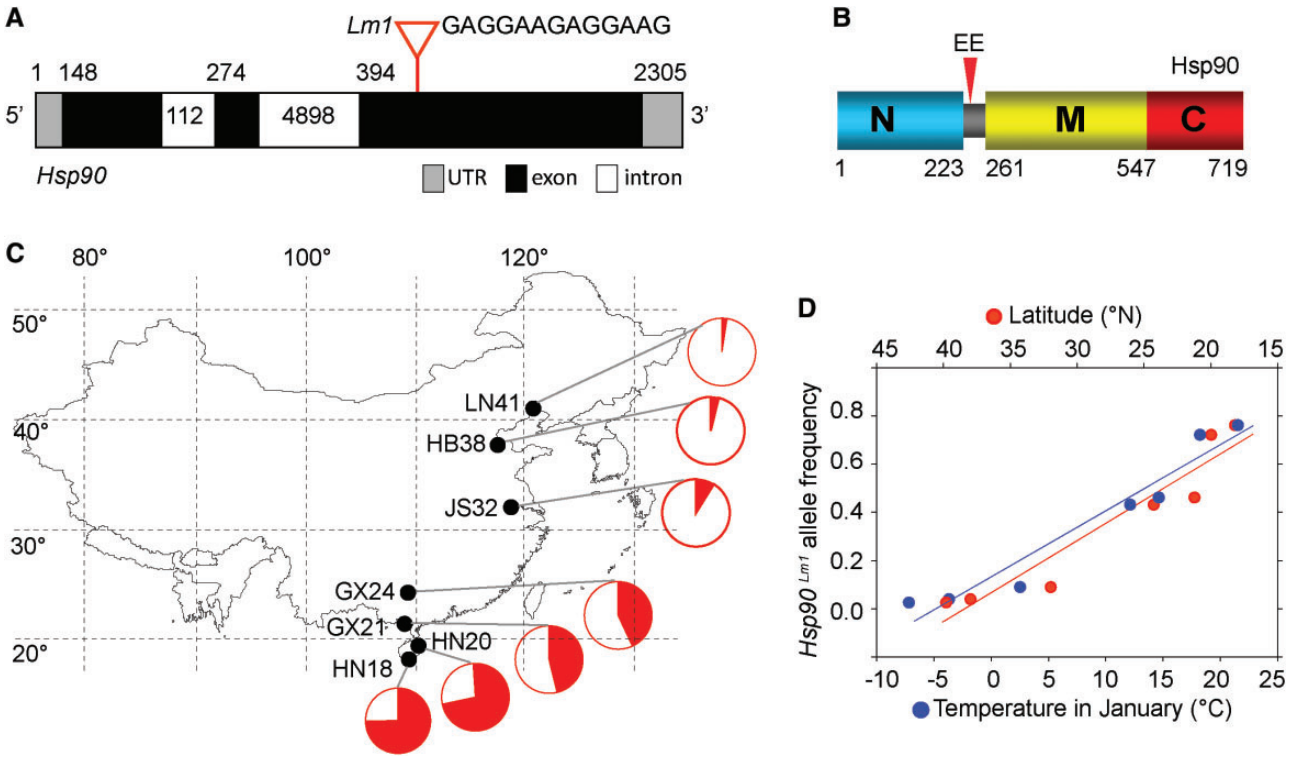

Fig. 2. *Lm1*-mediated *Hsp90* mutation in locust natural populations. (*A*) Gene structure of *Hsp90* with *Lm1* insertional mutation. Numbers above the gene represent starting position of each exon into which intron is not counted. The *Lm1* inserts at 830 bp. Numbers inside the box represent length (bp) of each intron. *Lm1* insertion leads to a direct repeat of 13-bp fragment GAGGAAGAGGAAG flanking *Lm1*. (*B*) Predicted domain structure of Hsp90. N, M, and C represent N-terminal domain, Middle segment and C-terminal domain. In the *Lm1*-containing alleles, the *Lm1* element is spliced out from *Hsp90* transcript and leaves with a deletion of GAAGAG encoding two glutamic acids (i.e., EE, at positions 227-228) as indicated by the triangle. (*C*) Frequency of $Hsp90^{Lm1/+}$ in locust natural populations. The pie chart shows the proportion of heterozygotes in each population indicated by filled cycle. Sample size: $n >= 36$. See the abbreviation for each locality in supplementary table S2, Supplementary Material online. (*D*) Correlation between $Hsp90^{Lm1}$ frequency and latitude/monthly temperature at the sampling localities. Regression line is fitted between $Hsp90^{Lm1}$ frequency vs. latitude (red) or local temperature in January (blue).

No homozygous mutation of *Hsp90* occurred in the natural populations. Only wild type ($Hsp90^{+/+}$) and heterozygote ($Hsp90^{Lm1/+}$) were found in wild locusts. In the three southern populations (HN18, HN19 and HN20), the heterozygotes are up to 76%, which significantly departs from Hardy–Weinberg Disequilibrium (HWD). No homozygous mutants were detected in locust populations from different latitudes and altitudes (see supplementary table S2, Supplementary Material online).

To determine if the homozygote is rare or present only under specific conditions, we reared the locust adults under different conditions and performed different crosses. The cross of heterozygotes predominantly produced heterozygotes (more than 92%), but no homozygous mutants were formed, regardless of different parental photoperiods and egg incubation temperatures (see supplementary table S5, Supplementary Material online). The observed genotype frequency significantly departs from the Mendel's law of segregation (Chi-square test, $P < 0.001$). The reciprocal crosses of heterozygote and wild-type locusts neither produced homozygous mutants. We also examined the genotype of their progenies at the stage of early egg, nymph, and adults as well as in unhatched eggs. No homozygous mutants were detected. The results imply a possible heterozygote advantage maintained by balancing selection at the *Hsp90* site.

Correlation between $Hsp90^{Lm1}$ Allele Frequency and Local Environments

The allele $Hsp90^{Lm1}$ in natural populations decreases its frequency with the latitude from south to north of China (fig. 2*C* and see supplementary table S2, Supplementary Material online). We plotted the frequency of $Hsp90^{Lm1}$ against latitude (fig. 2*D*). Regression analysis showed a significant latitudinal cline ($F_{1,6} = 48.8$, $P = 0.0009$, $R^2 = 0.95$). The mean temperature in January, that is, the lowest temperature of the year, decreases gradually with latitude (see supplementary table S2, Supplementary Material online). Regression with the temperature revealed a significant correlation between $Hsp90^{Lm1}$ frequency and low local temperature ($F_{1,6} = 76.6$, $P = 0.0003$, $R^2 = 0.94$) (fig. 2*D*). The results imply a close association of the occurring *Hsp90* mutation with local environmental adaptation.

We also investigated the possible influence of locust population structure on the correlation. The locust populations are diverged into the northern and southern lineage based on

mitochondrial genome sequences (Ma et al. 2012). We thus replotted the allele frequency against latitude and temperature of the localities within the northern and southern lineage, separately. In the southern lineage, the frequency of $Hsp90^{Lm1}$ is significantly correlated with latitude ($F_{1,6} = 16.3$, $P = 0.016$, $R^2 = 0.90$) and local temperature ($F_{1,6} = 14.8$, $P = 0.018$, $R^2 = 0.89$). In the northern lineage, the allele frequency is also significantly correlated with latitude ($F_{1,5} = 19.1$, $P = 0.022$, $R^2 = 0.93$) and temperature ($F_{1,5} = 154.8$, $P = 0.001$, $R^2 = 0.99$) (see supplementary table S2, Supplementary Material online). Thus, the results support a correlation between the *Hsp90* mutation and local environments within locust lineage, although the correlation is weaker than that across population structure.

Selective Sweep Associated with the *Lm1* Insertion in *Hsp90*

We then attempted to ask whether the *Lm1* insertion into *Hsp90* is beneficial (or linked to a beneficial mutation). If yes, we expected reduced variation at sites linked to the *Lm1*, which is a signature of selective sweep. To test this prediction, we sequenced the coding regions flanking the *Lm1* in 30 HN18 strains and 22 GX24 strains: one 350 bp immediately upstream and another 1.4 kilo-bp (kb) immediately downstream the *Lm1*. The haplotypes containing the *Lm1* are divergent from the haplotypes without the *Lm1* (fig. 3). The haplotype structure near the *Lm1* insertion site revealed a dramatically reduced sequence variability in the mutant alleles. We found only two distinct haplotypes among the 28 alleles containing the *Lm1*. In contrast, each one of the 24 wild-type alleles represents a unique haplotype (fig. 3). In 5×10^4 coalescence simulation, we failed to generate any samples with fewer than 10 distinct haplotypes (compared with the observed two distinct haplotypes, 95% confidential limit) linked to a polymorphism as frequent as the *Lm1* ($P < 0.000001$, assuming no recombination). This result supports the hypothesis of a recent incomplete selective sweep.

To test whether the selective sweep is completely linked to $Hsp90^{Lm1}$, we studied the polymorphism in the regions far from the insertion site (see supplementary fig. S2, Supplementary Material online). At a region ~ 7 kb upstream or ~ 5 kb downstream the *Lm1*, we found a modest signature of selection ($P = 0.057$ and 0.037 in 5×10^4 coalescence simulations, respectively). At a distance of 14 kb upstream or 15 kb downstream the *Lm1*, coalescent simulation revealed no violation of neutrality ($P > 0.05$ in each case). Therefore, the selective sweep decays at increasing distance from the *Lm1* insertion, suggesting a focal point of sweep in or close to the *Lm1* insertion.

We further performed McDonald-Kreitman test to look for evidence of selection in the evolution of this genomic region (Mcdonald and Kreitman 1991). The ratio of amino acid replacement to synonymous substitutions that are fixed between species is significantly lower than the ratio of replacement to synonymous polymorphisms (*G*-test of independence, $G = 39.1$, $P = 3.94E\text{-}10$; Fisher's exact test, $P = 3.84E\text{-}09$) (table 1). About 48% of the polymorphisms are replacement substitutions, but only one fixed differences between species are replacement substitutions. 29% of the polymorphisms are replacement substitutions after singletons are excluded. This observation is against the neutral expectation.

The excess amino acid replacement polymorphism predominantly occurred at two regions: one region contains four amino acid replacements at positions 203–223 in N-terminal domain of Hsp90, and another region contains six amino acid replacements at positions 401–452 in Middle domain. Among these replacements, two amino acid substitutions in the two loci (Ile→Phe 203, segregating sites -74 and Val→Ala 452, segregating site 673) are fixed in the derived states within the alleles containing the *Lm1* but fixed in the ancestral states within the alleles lacking the *Lm1* (fig. 3). We further sequenced the region flanking the *Lm1* insertion in another 14 strains of locusts from around the world (France, Kazakhstan, Mongolia, Japan, and different regions of China). The two novel amino acids are also fixed in the *Lm1*-bearing alleles (see supplementary table S6, Supplementary Material online). However, it cannot be determined at this moment which genetic change, the *Lm1* insertion or the two amino acid substitutions, was the first target of selection.

Photoperiod-Dependent Embryonic Developmental Variation Associated with *Hsp90* Mutation

We next ask what the beneficial effect the *Lm1*-containing allele might have. We postulated that the *Hsp90* mutation might affect embryonic development because Hsp90 can regulate the embryonic development in response to environmental cues in *L. migratoria* as well as in other insects (Bradley et al. 2012; Chen et al. 2015; King and MacRae 2015; Li 2016). We thus compared the embryonic developmental rate of $Hsp90^{+/+}$ and $Hsp90^{+/Lm1}$ eggs from HN18 isogenic lines whose parents were reared at three different photoperiods (Light vs. dark hours (L:D) 8:16, 12:12, and 16:8) (fig. 4). Generally, the embryonic developmental rate decreases with day length, but differs between the genotypes. At photoperiod LD8:16, the wild-type and heterozygous embryos developed at almost the same rate, to stage 23.8 and 23.7, respectively, after incubation at 30 °C for 9 days. However, the embryos developed to stage 22.1 ($Hsp90^{+/+}$) and stage 23.2 ($Hsp90^{+/Lm1}$) at LD12:12, indicating that the heterozygote developed faster than the wild type at this photoperiod (Mann–Whitney *U* test, $P < 0.001$). The heterozygous embryo also develops faster than the wild type at long day length (LD16:8) ($P < 0.001$), with stage 20.1 for the wild type and stage 22.6 for the heterozygote (fig. 4*A*). Therefore, the *Hsp90* mutation promotes embryonic development; this promotion is further enhanced by long-day light cues.

The *Hsp90* genotypes also differ in the within-population variation in embryonic development. We observed in the heterozygotes a statistically significant decrease in the relative standard deviation (RSD, −35% compared with the wild type; Levene's test, $P = 0.038$) at LD8:16. The same results were observed at LD12:12 (RSD, −61% compared with the wild type; $P = 0.001$) and at LD16:8 (RSD, −67% compared with the wild type; $P < 0.001$). Thus, the heterozygous embryos

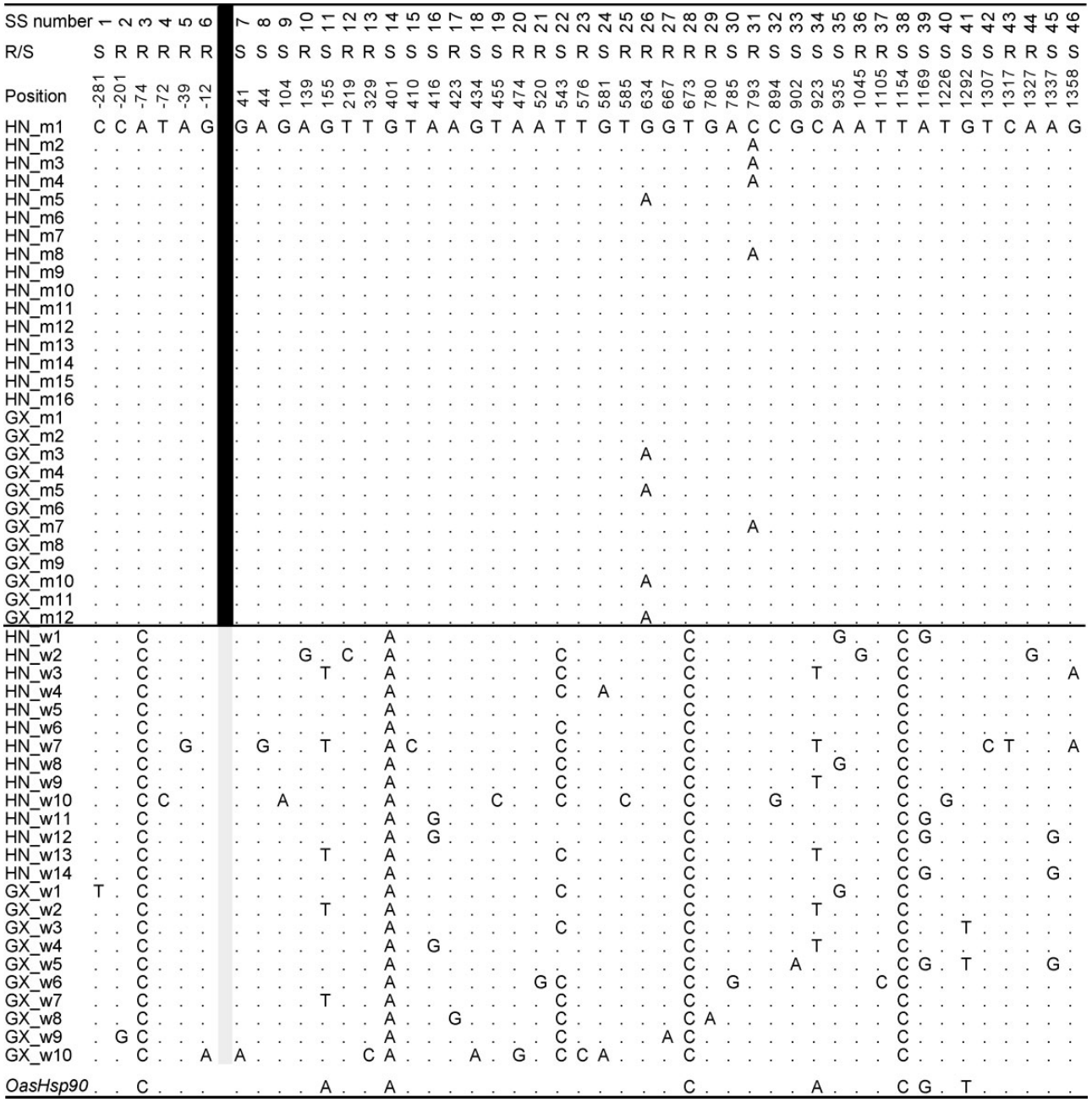

Fig. 3. Genetic variability in sites flanking *Lm1* insertion in *Hsp90* coding region. The segregating sites (SS) within the 1,762 bp region flanking the *Lm1* insertion are shown. The 28 sequences above the middle line are mutant, and the 24 sequences below the middle line are wild type. *Lm1* is shown as a black column; the absence of *Lm1* is shown as an empty column. The SS number and the distance from *Lm1* are at the top. The SS are represented as replacement (R) or synonymous (S) polymorphisms. The nucleotides of *Hsp90* in the outgroup *Oedaleus asiaticus* (*OasHsp90*) corresponding to the segregating sites are shown at the bottom.

exhibit less variation in the developmental trait, that is, higher developmental synchrony, than the wild type (fig. 4*A* as well as the inset figures).

To test whether the embryonic developmental variation is affected by genetic background of source populations, we examined the embryos in isogenic lines in three populations, HN18, HN19, and HN20, under long photoperiod (fig. 4*B* and see supplementary table S1, Supplementary Material online). We observed significant decrease in the variation of embryonic development of heterozygotes in HN18 (RSD, −52% compared with the wild type; Levene's test, $P < 0.001$), in HN19 (RSD, −37% compared with the wild type; $P < 0.001$), and in HN20 (RSD, −63% compared with the wild type; $P < 0.001$). The results reveal a clear divergence in embryonic developmental variability between the *Hsp90* genotypes independent of their origins.

In summary, the embryonic developmental rate decreased with day length of parental photoperiod. However, the heterozygous embryo develops faster than the wild type, prominently under long-day photoperiod. Meanwhile, the *Hsp90*

heterozygotes exhibit a reduced developmental variation and consequently a more synchronized development of embryos (fig. 4C).

Table 1. The McDonald-Kreitman Test of Selection on *Hsp90*.

	Fixed	Polymorphic	*P* value
Synonymous	**49**	**24 (12)**	**3.94E−10**
Replacement	**0**	**22 (5)**	**–**

NOTE.—Number of replacement and synonymous substitutions for fixed differences between species and polymorphisms within species of *L. migratoria* is provided. The fixed difference is calculated in comparison with the sequences of the grasshopper *O. asiaticus*. The 1.8 kb coding region of *Hsp90* was examined (see fig. 3). A G-test of independence (with the continuity correction) was used to test the null hypothesis that the ratio of replacement to synonymous fixed differences between species is same as the ratio of replacement to synonymous polymorphism within species. Values excluding singletons are given in parentheses.

Discussion

Contribution of *Lm1* Elements to Genomic Evolution

TEs were reported to cause adaptive genetic and genomic changes in some cases. However, the effect of TEs in the essential genes and mediation of phenotypic variation in adaptive evolution remains less understood. We characterized the families of *Lm1* elements in the locust genome and found that *Lm1* elements are highly abundant and ubiquitous in the locust genome. In addition to the locust, *Lm1*-like sequences are also found in other orthopteran species (Wu et al. 2001; Aguirre et al. 2010; Berthier et al. 2011), but not recorded in other insect orders. Thus, *Lm1* elements seem to be orthopteran-specific TEs. However, all *Lm1* sequences possess the two consensus elements (box A and B) that were presumably recruited from the internal promoter of Pol III

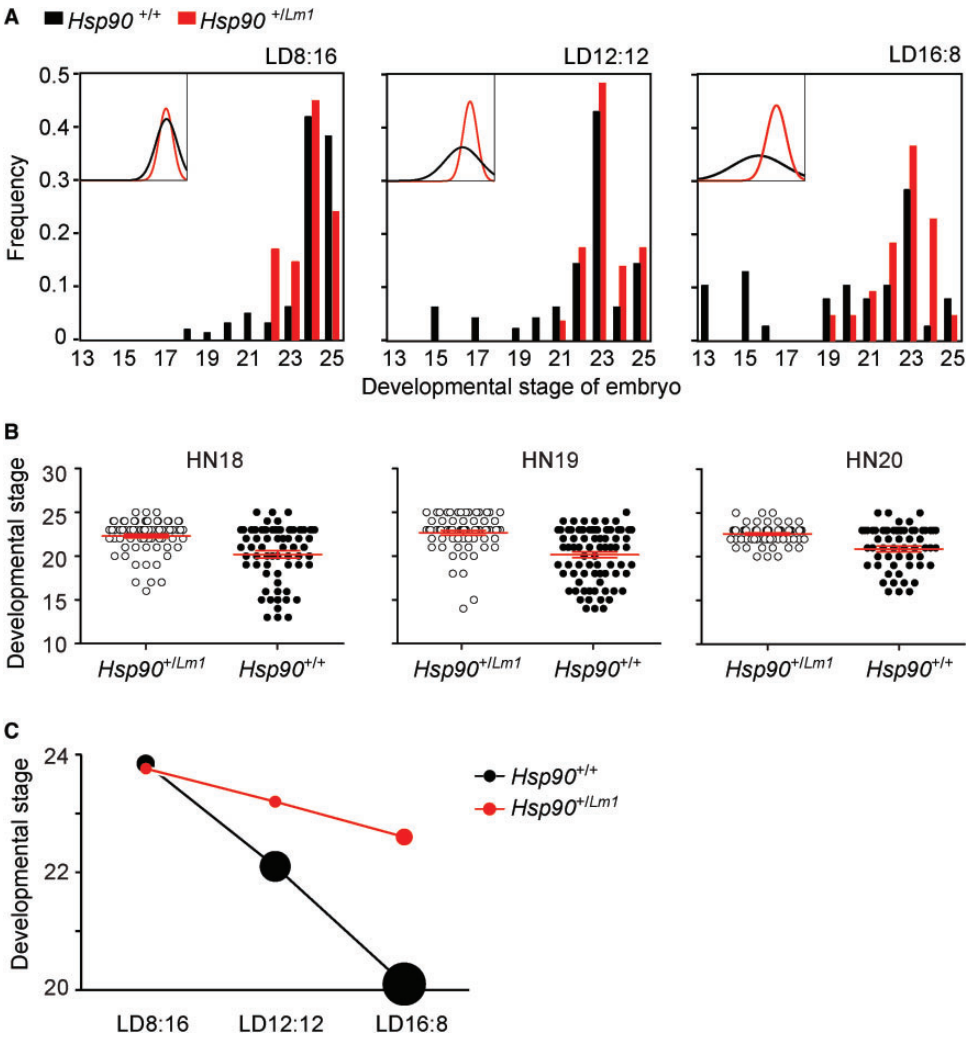

FIG. 4. Embryonic development of locusts in response to parental photoperiods. (*A*) Development of eggs produced by adults reared with three photoperiods (LD8:16, LD12:12, and LD16:8) at 30 °C. Eggs had been incubated at 30 °C for 9 days before examination of embryonic developmental stage. Black column, *Hsp90* wild type ($Hsp90^{+/+}$); Red column, *Hsp90* bearing *Lm1* insertion ($Hsp90^{Lm1/+}$). Inset figure shows the normal distribution of the developmental stages. Eggs from ten isofemale strains were examined. $N = 165$ and 125 (LD8:16), 147 and 142 (LD12:12), 117 and 131 (LD16:8), for $Hsp90^{+/+}$ and $Hsp90^{Lm1/+}$. (*B*) Variation in embryonic developmental stages in three populations. Parent adults from the natural populations (HN18, HN19, and HN20) were reared at long photoperiods (LD16:8) at 30 °C. Eggs from ten isofemale strains were examined. $N = 63$ and 70 (HN18), 75 and 68 (HN19), 57 and 69 (HN20) for $Hsp90^{+/+}$ and $Hsp90^{Lm1/+}$, respectively. See the population information in supplementary table S2, Supplementary Material online. The mean ± 1 SE are shown (red). (*C*) Schematic representation of embryonic developmental changes with parental photoperiods. The cycle size dictates relative standard deviation.

from 7SL or transfer RNAs (Van Arsdell et al. 1981; Labuda et al. 1991). Mammalian ubiquitous SINEs (e.g., human *Alu* family) also harbor the bipartite consensus elements (Labuda et al. 1991). Therefore, *Lm1* elements could be the counterparts of the ubiquitous SINE families that widely occur in eukaryotes.

The *Lm1* elements have proliferated to millions of copies after their invasion to the locust genome, comprising 2.2% of the locust genome. The efficient proliferation of *Lm1* elements may contribute to their structural attributes as retroposons. The two RNA Pol III promoter sequences could direct their own transcription and facilitate their access to retropositional machinery (Labuda et al. 1991; Gilbert and Labuda 1999).

The *Lm1* elements have caused the high genomic diversity. The majority of *Lm1* sequences (91%) are interspersed in the intergenic regions, and a few of the *Lm1* elements are transcribed with mRNA, for example, in the hunchback transcription factor (Patel et al. 2001), and the juvenile hormone-regulated gene (Zhang and Wyatt 1996). *Lm1* thus could contribute to the regulatory changes as well as the origin of novel transcripts and even proteins. For example, the *Lm1* insertion in *Hsp90* leads to a deletion of two amino acids. *Lm1* is also possibly associated with DNA methylation, and thus, gene expression in the locusts (Robinson et al. 2016). We observed high proportion of *Lm1* elements (20 in 24) exhibiting insertion polymorphism in wild locusts. Therefore, *Lm1* could drive the genomic evolution and regulatory changes in natural populations.

Heterozygote Advantage and Balancing Selection at *Hsp90*

The TE insertion event in *Hsp90* causes both genetic and possibly biochemical changes associated with Hsp90. The *Lm1* element in the mutant allele is spliced out possibly through the transposition machinery associated with its structural attributes (Gilbert and Labuda 1999), or the introduction of alternative splicing sites with the insertion (Sorek et al. 2002), or other unknown mechanisms. Similar phenomena were reported with other TEs (Wessler et al. 1987; Sorek et al. 2002). The two-amino acid deletion caused by the *Lm1* splicing is present at the linker region of Hsp90 located between N-terminal domain and middle segment. Different from other Hsp90 domains, this region is poorly conserved and absent from the bacterial HtpG and mitochondrial proteins (Pearl and Prodromou 2006). Removal of this region does not noticeably impair the essential functions of yeast Hsp90 in vivo (Scheibel et al. 1999; Jahn et al. 2014). However, functional involvement by this segment is suggested, for example, modulating domain contacts and Hsp90 chaperone activity (Tsutsumi et al. 2012; Jahn et al. 2014). Therefore, this specific insertional mutation in *Hsp90* may be consequentially mild but significant. In contrast, insertional mutations in other regions of *Hsp90* can be lethal, or incur high fitness costs in organisms (Rutherford and Lindquist 1998; Chen and Wagner 2012). Meanwhile, natural selection has acted on the region centered at the *Lm1* element, resulting to excess amino acid replacement polymorphisms in N-terminal domain and the middle domain of Hsp90. The polymorphisms are essential for ATP and client protein binding. Among these replacements, two amino acid substitutions are fixed in the derived states. The three major genetic changes in *Hsp90*, i.e., the *Lm1* insertion and the two amino acid changes, could modulate Hsp90 chaperone activity and phenotypic variations.

The predominant presence of *Hsp90* heterozygote in the locust could be associated with the fundamental role of Hsp90 in cellular activity. Hsp90 is required for activating many signaling proteins and highly critical for morphogenesis and development; any mutation in either regulatory or coding region of *Hsp90* gene is lethal to *D. melanogaster* flies (Rutherford and Lindquist 1998; Chen and Wagner 2012). Even point mutations in *Hsp90* are homozygous lethal for *D. melanogaster* (Yue et al. 1999). Therefore, the heterozygous mutation at the linker region of *Hsp90* may cause less harmful or neutral effect on organisms, and instead have conferred the developmental advantages under specific environmental conditions (see more discussion below).

Balancing selection could have worked to maintain the heterozygote advantage. The southern-most populations contain more than 71% locusts bearing $Hsp90^{Lm1}$ allele. In contrast, the northern-most populations harbor much less of the mutant allele. This finding is the first discovery of naturally occurring and consequential *Hsp90* mutant that has swept high frequency in the wild. The locust is a highly migratory species with populations breaking out frequently in south and north of China (Jing and Kang 2003; Ma et al. 2012). Hence, the locusts are usually expected to have minimal population substructure and strong gene flow counteracting random drift (Ma et al. 2012; Chapuis et al. 2014). Latitude associated local environments have been considered to exert selective pressure and contribute to the genetic differentiation of the locust populations (Wang and Kang 2005; Zhang et al. 2009; Ma et al. 2012). The cline of $Hsp90^{Lm1}$ allele is correlated with latitude and local environments, implying a possible role of this mutation in local adaptation. It is believed that genetic diversity would be severely reduced in haplotypes carrying that advantageous allele when the allele is maintained by balancing selection and does not become fixed in the population (Charlesworth 2006). We indeed observed a strong selective sweep linked to the *Lm1* insertion. Selection would purge other *Hsp90* variants from natural populations. This variant is beneficial, and therefore persists and accumulates in locust populations.

Beneficial Consequences in Embryonic Development and Seasonal Occurrence

The *Hsp90* mutation is consequential for embryonic development, and could contribute to the evolution of voltinism in the locust. In general, the developmental rate of insects decreases with day length, or increases at short-day seasons (Aldyhim and Khalil 1993; Leimar 1996). The same phenomenon was observed in the *Hsp90* wild-type and heterozygous embryos. The short-day photoperiod in autumn could promote rapid embryonic development in preparation for overwintering. However, the two genotypes respond differently in

embryonic development to parental photoperiod regime. The heterozygous embryos develop faster than the wild type, particularly at the long-day photoperiod. The long-day period represents the natural light condition in north hemisphere during vernal and autumnal equinox. Thus, the heterozygotes predominantly occurring in the southern populations will develop fast and thereby produce multiple generations in the growing season (i.e., multivoltine). In fact, a latitudinal cline of locust voltinism exists: locusts in the most northern population (e.g., LN41 and HB38) are obligate univoltine; locusts in central China (e.g., JS32 and GX24) are facultative or bivoltine; and locusts in the most southern populations (e.g., HN18, HN19 and HN20) are multivoltine (Ma 1958; Jing and Kang 2003). Therefore, the *Lm1*-mediated mutation may be behind the evolution of locust voltinism and thus promotes locusts' adaptation to seasonal environments. Meanwhile, photoperiod change is indicative of seasonal temperature, with which the allele frequency of the mutation in latitudinal locust populations is correlated (Fig. 2). The cold tolerance of the locust eggs presents a similar latitudinal cline (Jing and Kang 2003). Thus, the *Hsp90* mutation may also confer adaptation of the locust to local temperature.

Our study also implicates an evolutionary mechanism for the regulation of developmental synchrony by Hsp90. The *Hsp90* mutant alleles have been subjected to selective sweep that results into a reduced genetic variation in genomic region flanking the *Lm1* insertion. Meanwhile, the mutant lines present a reduced variation in embryonic development, that is, high developmental synchrony. In contrast, the wild-type locusts exhibit greater developmental variation. In other words, the extent of developmental variation is related to the *Lm1*-linked polymorphic state in *Hsp90* gene region. We propose that the genetic homogenization in the mutant alleles could promote the developmental synchronization. The developmental synchrony could benefit the multivoltinism of the locusts in southern populations by facilitating the locusts' mating, reproduction and migration so that they can complete multiple generations in a season.

Hsp90 can regulate embryonic development in response to photoperiod cues through several mechanisms. Hsp90 utilizes the energy of ATP hydrolysis for binding client proteins and maintaining proteins in an active conformation (Echeverria and Picard 2010). The range of Hsp90 client proteins includes steroid hormone receptors, proteins involved in signal transduction pathways, and several transcription factors. In addition to the cytosolic function, Hsp90 can control gene expression and alter cellular physiology during development by globally targeting the RNA polymerase II complex at the promoter of numerous genes (Sawarkar et al. 2012). The target genes include transcription factors, signaling molecules, Hox genes, and environmental-responsive genes that can sense the environmental change and regulate development. Thus, the embryonic development could be regulated by photoperiod signals through induction of a few genes that require functional Hsp90 for their optimal induction. In addition, Hsp90 buffers the expression of cryptic genetic variation that affects phenotypic variation. Hsp90 inhibition or gene mutation alters its buffering capacity, and possibly promotes the expression of novel traits (Rutherford and Lindquist 1998; Rutherford et al. 2007; Chen and Wagner 2012; Rohner et al. 2013). Therefore, the *Lm1*-mediated *Hsp90* mutations could control the expression of the embryonic developmental variation.

Materials and Methods

Locusts

At least 800 wild locusts were collected in field in each locality (see supplementary table S1, Supplementary Material online). These collection places are located along eastern coastal area of China where locust plagues had historically occurred frequently (Ma 1958). The locusts of each population were sampled at more than 10 sites with an interval of $>$ 500 m. The live locusts collected in Hainan province (HN18) and Hebei province (HB38) were retrieved and maintained in a large stock ($>$ 800 locusts) in the laboratory of the Institute of Zoology, Chinese Academy of Sciences in Beijing.

Locust colonies were reared in well-ventilated cages ($35 \times 35 \times 35$ cm) at a density of approximately 400 individuals per cage. Locusts were fed with fresh wheat seedlings and bran. The culturing environment was kept constant in an LD14:10 photoperiod regime at $30 \pm 2\,^{\circ}$C. Newly laid eggs were kept in a plastic cup (diameter of 6 cm and height of 9 cm) filled with sterilized sands with 8% humidity and maintained in an incubator at $30 \pm 1\,^{\circ}$C.

Sequence Analysis

Lm1 consensus sequence of the locust was initially obtained from NCBI database (Bradfield et al. 1985) and validated in the locust by sequencing. The *Lm1* sequence was blasted against the locust genome to search for *Lm1*-like sequences (BLAST E value: $<$ 1e-20) (Wang et al. 2014). All *Lm1*-like sequences were clustered using the program CD-hit to make non-redundant *Lm1* sequence databases (Li and Godzik 2006). It generates a database of representative sequences for all clusters. The core SINE sequence was obtained from a previous report (Gilbert and Labuda 1999).

The evolutionary tree of *Lm1* elements was constructed using the Maximum Likelihood method that is based on the representative sequences of the first 1,000 largest *Lm1* clusters. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013). The domain structure of Hsp90 was predicted on protein sequence conservation with yeast Hsp90 (Pearl and Prodromou 2006). Coalescence simulations were performed using DnaSP5.10 (Librado and Rozas 2009).

Lm1 Insertion Polymorphism and Genotyping

We examined the *Lm1* insertion polymorphisms, that is, polymorphisms with the presence/absence of an *Lm1* element at a particular chromosomal location, by polymerase chain reaction (PCR) in natural populations. Based on the reference locust genome (Wang et al. 2014) we investigated 24 gene regions in which *Lm1* element of a specific clade (cluster 324, see supplementary table S1, Supplementary Material online)

is present. A total of 30 individuals from each of six latitudinal populations were genotyped. The six populations are HN18, GX21, GX24, JS32, HB38, and LN41, which are located between latitudes 18°N and 41°N (see supplementary table S2, Supplementary Material online). Genomic DNA from all these locusts was extracted using DNeasy Tissue kit (Qiagen). To assay for *Lm1* insertion, we designed one set of primers for each loci, which lay in the flanking region approximately 200 bp to the left or right of the *Lm1* insertion (see supplementary table S7, Supplementary Material online). We expected this PCR to present a higher band when the element is present. The presence of *Lm1* was further confirmed by sequencing of the PCR product.

We screened four *Hsp* genes, i.e., *Hsp90*, *Hsp70*, *Hsp40*, and *Hsp20.5*, by PCR for *Lm1* insertion in their coding and 5′ proximal promoter region. The sequences of the four *Hsp* genes were previously reported (Wang et al. 2007; Wang et al. 2014). To detect the presence of *Lm1* element, we designed two sets of primers that consist of a "left" primer complementary to the *Hsp* gene and a "right" primer complementary to the *Lm1* sequence, or in reverse direction. PCR products were sequenced to confirm the specificity of amplification.

Two PCRs were performed to detect the specific *Lm1* insertion polymorphism in the *Hsp90*. In one PCR, genomic region covering the *Lm1* insertion site was amplified using two *Hsp90*-specific primers. In another PCR, presence of *Lm1* insertion was detected with one primer specific to *Hsp90* sequence and another primer specific to *Lm1* sequence. Amplification with the heterozygote DNA is expected to give two bands in the first PCR and one band in the second PCR. The *Hsp90*-specific primers blasted against the locust genome and transcriptomes to confirm the specificity of amplification for these primers. Primers for the PCRs are listed in supplementary table S7, Supplementary Material online.

SNP Sequencing in *Hsp90* and Flanking Regions

We sequenced a total of ~1.8 kb of *Hsp90* coding region flanking the *Lm1* insertion, including 350 bp immediately upstream and 1.4 kb immediately downstream of *Lm1*. The sequenced samples were from 30 isofemale lines established from the HN18 population and 22 isofemale lines from GX21 population. The nucleotides of *Hsp90* (*OasHsp90*) in an outgroup grasshopper species *O. asiaticus* were also sequenced. *L. migratoria* belongs to the monotypic genera *Locusta*, and *O. asiaticus* is an Orthopteran species phylogenetically most closely related to *L. migratoria* (Fries et al. 2007; Ma et al. 2009). Five *O. asiaticus* strains were sequenced. The wild-type sequences in *L. migratory* were amplified with *Hsp90*-specific primers. The mutant sequences were amplified with one primer *Lm1*-specific and another *Hsp90*-specific to obtain the region upstream and downstream of the *Lm1* insertion in two separate PCRs. The final PCR product was purified and sequenced from two ends. Primers for the PCRs are listed in supplementary table S7, Supplementary Material online. All the sequences have been deposited in GenBank under accession numbers KX982265–KX982316 (upstream sequences) and KY022100–KY022151 (downstream sequences).

We sequenced four single-copy intergenic regions at various distances from the *Lm1* insertion. The four regions are 1) 490 bp in length, ~7 kb upstream the *Lm1*, 2) 479 bp, ~14 kb upstream, 3) 501 bp, ~5 kb downstream, and 4) 494 bp, ~15 kb downstream. To obtain the specific sequence of these fragments located on the strand bearing the *Lm1* in the heterozygote, we performed three nested long PCRs involving three sets of primers. In the first PCR run, the DNA fragment covering the *Lm1* and the target sequence was amplified with one primer specific to *Hsp90* and another specific to the target sequence. In the second PCR run, PCR was performed with one primer specific to *Lm1* and another specific to the target sequence to obtain the fragment bearing the *Lm1*. In the third PCR run, PCR was performed with two primers specific to the target sequence. The PCR product in the first two PCRs was purified and used as PCR template in the following PCR. To obtain the four sequences in the wild-type locusts, we only performed the third PCR used with the target sequence-specific primers. The final PCR product was purified and sequenced from two ends. We also confirmed with *Lm1*-specific primers that there is no other *Lm1* present in these regions in the strains we screened. The LA PCR™ Kit (Takara Bio) was used for the long range PCR according to the user's manual. A total of 25 locusts were analyzed, including 15 mutant locusts and 10 wild-type locusts from the HN18 and GX21 population. Primers for the PCRs are listed in supplementary table S7, Supplementary Material online. All the sequences have been deposited in GenBank under accession numbers KY022000–KY022099.

Embryonic Development

Locust isofemale lines used in phenotypic measurement were established from HN18 population. The $Hsp90^{+/+}$ and $Hsp90^{+/Lm1}$ adults from an isofemale line were mated to produce eggs that were examined for embryonic development and genotyping. To set up different photoperiod regimes (L:D 8:16, 12:12 and 16:8), locusts were reared from 5th-instar nymphs in a well-ventilated incubator in which light and temperature were programmably controlled. The temperature was set at 30 ± 1 °C. The parent locusts (nymphs or adults) were genotyped using genomic DNA extracted from 1 μL hemolymph (Holehouse et al. 2003). Approximately, 30 eggs from each of 10 isofemale lines were examined. When standard incubated eggs developed into the 9-day-old stage, embryonic development was examined under a microscope (Leica Application suite M205C, v. 3.3.0, Wetzlar Germany). The whole developmental period of the locust embryos was classified into 27 distinct developmental stages according to the quantitative staging system previously described (Bentley et al. 1979). The eggs were then stored in 70% ethanol for DNA purification and genotyping.

Statistics

The departure of allele frequency from HWD was analyzed using Chi-square test. The regression analysis of latitudinal clines included seven localities (e.g., HN18, HN20, GX21,

GX24, JS32, HB38, and LN41) while not considering population structure, and included six localities (e.g., HN18, HN19, HN20, GD21, GX21, and GX24) for the southern lineage, and five localities (e.g., JS32, AH33, HN34, HB38, and LN41) for northern lineage (see supplementary table S2, Supplementary Material online). The frequency distribution of the developmental stages was evaluated using the Mann–Whitney *U* test. The normal distribution of developmental stages was approximated through kernel density smoothing using the kernel function $K(t) = (2\pi)^{-1/2}\exp(-t^2/2)$ (Wand and Jones 1995). Coefficient of variation of developmental rate, also known as RSD, is defined as the ratio of standard deviation to mean. Difference in RSD was analyzed using Levene's test. McDonald-Kreitman test was performed to look for evidence of positive selection (Mcdonald and Kreitman 1991). The test was used to test the null hypothesis that the proportion of replacement substitutions is independent of whether the substitutions are fixed or polymorphic, based on a neutral model of evolution. Replacement refers to nucleotide differences within codons that alter the amino acid (Sawyer and Hartl 1992). Differences were considered statistically significant if $P < 0.05$. Data were analyzed using the SPSS 16.0 software (SPSS, Chicago, IL, USA).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Shuguang Hao for providing *O. asiaticus* sample and statistical analysis, Liang Leng, Yi Zhang and Gongxin Peng for help with bioinformatic analysis, and Arnar Pálsson, Menghua Li, two reviewers and the Associate Editor for helpful comments on the manuscript. This work was supported by the National Natural Science Foundation of China (grant number 91331106, 31472048, and 31172148). All the sequences have been deposited in GenBank under accession numbers KX982265–KX982316, KY022100–KY022151, and KY022000–KY022099.

References

Ågren J, Wright S. 2011. Co-evolution between transposable elements and their hosts: a major factor in genome size evolution? *Chrom Res*. 19:777–786.

Aguirre MP, Bloor P, Ramírez-Escobar U, Ortego J, Cordero PJ. 2010. Isolation and characterization of polymorphic microsatellite markers in the grasshopper *Mioscirtus wagneri* (Orthoptera: Acrididae). *Conserv Genet*. 11:1119–1121.

Aldyhim YN, Khalil AF. 1993. Influence of temperature and daylength on population development of Aphis gossypii on Cucurbita pepo. *Entomol Exp Appl*. 67:167–172.

Aminetzach YT, Macpherson JM, Petrov DA. 2005. Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila*. *Science* 309:764–767.

Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. *Nat Rev Gene*. 3:370–379.

Begun DJ, Kofler R, Betancourt AJ, Schlötterer C. 2012. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genet*. 8:e1002487.

Bentley D, Keshishian H, Shankland M, Toroian-Raymond A. 1979. Quantitative staging of embryonic development of the grasshopper, *Schistocerca nitens*. *J Embryol Exp Morph*. 54:47–74.

Berthier K, Chapuis M-P, Moosavi SM, Tohidi-Esfahani D, Sword GA. 2011. Nuclear insertions and heteroplasmy of mitochondrial DNA as two sources of intra-individual genomic variation in grasshoppers. *Syst Entomol*. 36:285–299.

Biemont C, Vieira C. 2006. Genetics: junk DNA as an evolutionary force. *Nature* 443:521–524.

Bishop CD, Bates WR, Brandhorst BP. 2002. HSP90 function is required for morphogenesis in ascidian and echinoid embryos. *Dev Genes Evol*. 212:70–80.

Borkovich KA, Farrelly FW, Finkelstein DB, Taulien J, Lindquist S. 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.

Bradfield JY, Locke J, Wyatt GR. 1985. An ubiquitous interspersed DNA sequence family in an insect. *DNA* 4:357–363.

Bradley E, Bieberich E, Mivechi NF, Tangpisuthipongsa D, Wang G. 2012. Regulation of embryonic stem cell pluripotency by heat shock protein 90. *Stem Cells*. 30:1624–1633.

Britten RJ. 2010. Transposable element insertions have strongly affected human evolution. *Proc Natl Acad Sci U S A*. 107:19945–19948.

Chapuis MP, Plantamp C, Blondin L, Pages C, Vassal JM, Lecoq M. 2014. Demographic processes shaping genetic variation of the solitarious phase of the desert locust. *Mol Ecol*. 23:1749–1763.

Charlesworth B, Langley CH. 1989. The population genetics of *Drosophila* transposable elements. *Annu Rev Genet*. 23:251–287.

Charlesworth D. 2006. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet*. 2:e64.

Chen B, Li S, Ren Q, Tong X, Zhang X, Kang L. 2015. Paternal epigenetic effects of population density on locust phase-related characteristics associated with heat-shock protein expression. *Mol Ecol*. 24:851–862.

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.

Chen B, Walser J-C, Rodgers T, Sobota R, Burke M, Rose M, Feder M. 2007. Abundant, diverse, and consequential P element segregate in promoters of small heat-shock genes in *Drosophila* populations. *J Evol Biol*. 20:2056–2066.

de Koning AJ, Gu W, Castoe TA, Batzer MA, Pollock DD. 2011. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet*. 7:e1002384.

Ding Y, Berrocal A, Morita T, Longden KD, Stern DL. 2016. Natural courtship song variation caused by an intronic retroelement in an ion channel gene. *Nature* 536:329–332.

Echeverria PC, Picard D. 2010. Molecular chaperones, essential partners of steroid hormone receptors for activity and mobility. *Biochim Biophysic Acta (BBA)–Mol Cell Res*. 1803:641–649.

Ellison CE, Bachtrog D. 2013. Dosage compensation viatTransposable element mediated rewiring of a regulatory network. *Science* 342:846–850.

Ewing AD, Kazazian HH Jr. 2010. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res*. 20:1262–1270.

Feder ME, Hofmann GE. 1999. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu Rev Physiol*. 61:243–282.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet*. 9:397–405.

Fries M, Chapco W, Contreras D. 2007. A molecular phylogenetic analysis of the Oedipodinae and their intercontinental relationships. *J Orthopt Res*. 16:115–125.

Gilbert N, Labuda D. 1999. CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. *Proc Natl Acad Sci U S A*. 96:2869–2874.

Gonzalez J, Karasov TL, Messer PW, Petrov DA. 2010. Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. *PLoS Genet*. 6:e1000905.

Gonzalez J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. 2008. High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biol*. 6:e251.

Gonzalez J, Macpherson JM, Petrov DA. 2009. A recent adaptive transposable element insertion near highly conserved developmental loci in *Drosophila melanogaster*. *Mol Biol Evol*. 26:1949–1961.

Guio L, Barron MG, Gonzalez J. 2014. The transposable element Bari-Jheh mediates oxidative stress response in *Drosophila*. *Mol Ecol*. 23:2020–2030.

Holehouse KA, Hammond RL, Bourke AFG. (ref1 co-authors). 2003. Non-lethal sampling of DNA from bumble bees for conservation genetics. *Insectes Sociaux*. 50:277–285.

Huang LH, Wang HS, Kang L. 2008. Different evolutionary lineages of large and small heat shock proteins in eukaryotes. *Cell Res*. 18:1074–1076.

Jahn M, Rehn A, Pelz B, Hellenkamp B, Richter K, Rief M, Buchner J, Hugel T. 2014. The charged linker of the molecular chaperone Hsp90 modulates domain contacts and biological function. *Proc Natl Acad Sci U S A*. 111:17881–17886.

Jing X, Kang L. 2003. Geographical variation in egg cold hardiness: a study on the adaptation strategies of the migratory locust *Locusta migratoria* L. *Ecol Entomol*. 28:151–158.

Kang L, Chen B, Wei JN, Liu TX. 2009. Roles of thermal adaptation and chemical ecology in Liriomyza distribution and control. *Annu Rev Entomol*. 54:127–145.

King AM, MacRae TH. 2015. Insect heat shock proteins during stress and diapause. *Ann Rev Entomol*. 60:59–75.

Labuda D, Sinnett D, Richer C, Deragon J-M, Striker G. (Labuda1991 co-authors). 1991. Evolution of mouse B1 repeats: 7SL RNA folding pattern conserved. *J Mol Evol*. 32:405–414.

Leimar O. 1996. Life history plasticity: influence of photoperiod on growth and development in the common blue butterfly. *Oikos* 76:228–234.

Li H. 2016. Hsp90 confers the embryonic developmental robustness of offspring in the locust Locusta migratoria. Beijing: Institute of Zoology, Chinese Academy of Sciences.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.

Ma C, Liu C, Yang P, Kang L. 2009. The complete mitochondrial genomes of two band-winged grasshoppers, Gastrimargus marmoratus and Oedaleus asiaticus. *BMC Genom*. 10:156.

Ma C, Yang P, Jiang F, Chapuis MP, Shali Y, Sword GA, Kang L. 2012. Mitochondrial genomes reveal the global phylogeography and dispersal routes of the migratory locust. *Mol Ecol*. 21:4344–4358.

Ma SJ. 1958. The population dynamics of the oriental migratory locust (*Locusta migratoria* manilensis Meyen) in China. *Acta Entomol Sinica*. 8:1–40.

Mateo L, Ullastres A, Gonzalez J. 2014. A transposable element insertion confers xenobiotic resistance in *Drosophila*. *PLoS Genet*. 10:e1004560.

Mcdonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654.

Merenciano M, Ullastres A, de Cara MA, Barron MG, Gonzalez J. 2016. Multiple independent retroelement insertions in the promoter of a stress response gene have variable molecular and functional effects in *Drosophila*. *PLoS Genet*. 12:e1006249.

Patel NH, Hayward DC, Lall S, Pirkl NR, DiPietro D, Ball EE. 2001. Grasshopper hunchback expression reveals conserved and novel aspects of axis formation and segmentation. *Development* 128:3459–3472.

Pearl LH, Prodromou C. 2006. Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annu Rev Biochem*. 75:271–294.

Petrov DA, Fiston-Lavier AS, Lipatov M, Lenkov K, Gonzalez J. 2011. Population genomics of transposable elements in *Drosophila melanogaster*. *Mol Biol Evol*. 28:1633–1644.

Rebollo R, Romanish MT, Mager DL. 2012. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Ann Rev Genet*. 46:21–42.

Robinson KL, Tohidi-Esfahani D, Ponton F, Simpson SJ, Sword GA, Lo N. 2016. Alternative migratory locust phenotypes are associated with differences in the expression of genes encoding the methylation machinery. *Insect Mol Biol*. 25:105–115.

Rohner N, Jarosz DF, Kowalko JE, Yoshizawa M, Jeffery WR, Borowsky RL, Lindquist S, Tabin CJ. 2013. Cryptic variation in morphological evolution: HSP90 as a capacitor for loss of eyes in cavefish. *Science* 342:1372–1375.

Rutherford S, Hirate Y, Swalla BJ. 2007. The Hsp90 capacitor, developmental remodeling, and evolution: the robustness of gene networks and the curious evolvability of metamorphosis. *Critic Rev Biochem Mol Biol*. 42:355–372.

Rutherford SL, Lindquist S. 1998. Hsp90 as a capacitor for morphological evolution. *Nature* 396:336–342.

Sawarkar R, Sievers C, Paro R. 2012. Hsp90 globally targets paused RNA polymerase to regulate gene expression in response to environmental stimuli. *Cell* 149:807–818.

Sawyer SA, Hartl DL. 1992. Population genetics of polymorphism and divergence. *Genetics* 132:1161–1176.

Scheibel T, Siegmund HI, Jaenicke R, Ganz P, Lilie H, Buchner J. 1999. The charged region of Hsp90 modulates the function of the N-terminal domain. *Proc Natl Acad Sci U S A*. 96:1297–1302.

Schmidt JM, Good RT, Appleton B, Sherrard J, Raymant GC, Bogwitz MR, Martin J, Daborn PJ, Goddard ME, Batterham P, et al. 2010. Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus. *PLoS Genet*. 6:e1000998.

Sessegolo C, Burlet N, Haudry A. 2016. Strong phylogenetic inertia on genome size and transposable element content among 26 species of flies. *Biol Lett*. 12:20160407.

Sgrò CM, Milton CC, Jensen LT, Frydenberg J, Loeschcke V, Batterham P, Hoffmann AA. 2008. Nucleotide diversity in the Hsp90 gene in natural populations of *Drosophila melanogaster* from Australia. *Insect Mol Biol*. 17:685–697.

Sollars V, Lu X, Xiao L, Wang X, Garfinkel MD, Ruden DM. 2003. Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. *Nat Genet*. 33:70–74.

Song Y, Fee L, Lee TH, Wharton RP. 2007. The molecular chaperone Hsp90 is required for mRNA localization in *Drosophila melanogaster* embryos. *Genetics* 176:2213–2222.

Sorek R, Ast G, Graur D. 2002. Alu-containing exons are alternatively spliced. *Genome Res*. 12:1060–1067.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 30:2725–2729.

Tsutsumi S, Mollapour M, Prodromou C, Lee CT, Panaretou B, Yoshida S, Mayer MP, Neckers LM. 2012. Charged linker sequence modulates eukaryotic heat shock protein 90 (Hsp90) chaperone activity. *Proc Natl Acad Sci U S A*. 109:2937–2942.

Ullastres A, Petit N, Gonzalez J. 2015. Exploring the phenotypic space and the evolutionary history of a natural mutation in *Drosophila melanogaster*. *Mol Biol Evol*. 32:1800–1814.

van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. 2016. The industrial melanism mutation in British peppered moths is a transposable element. *Nature* 534:102–105.

Van Arsdell SW, Denison RA, Bernstein LB, Weiner AM, Manser T, Gesteland RF. 1981. Direct repeats flank three small nuclear RNA pseudogenes in the human genome. *Cell* 26:11–17.

Walser JC, Chen B, Feder ME. 2006. Heat-shock promoters: targets for evolution by *P* transposable elements in *Drosophila*. *PLoS Genet*. 2:e165.

Wand MP, Jones MC. 1995. Kernel smoothing. London: Chapman & Hall/CRC.

Wang HS, Wang XH, Zhou CS, Huang LH, Zhang SF, Guo W, Kang L. 2007. cDNA cloning of heat shock proteins and their expression in the two phases of the migratory locust. *Insect Mol Biol*. 16:207–219.

Wang X, Fang X, Yang P, Jiang X, Jiang F, Zhao D, Li B, Cui F, Wei J, Ma C, et al. 2014. The locust genome provides insight into swarm formation and long-distance flight. *Nat Commun*. 5:2957.

Wang XH, Kang L. 2005. Differences in egg thermotolerance between tropical and temperate populations of the migratory locust *Locusta migratoria* (Orthoptera: Acridiidae). *J Insect Physiol*. 51:1277–1285.

Wessler Baran G, Varagona M. 1987. The maize transposable element Ds is spliced from RNA. *Science* 237:916–918.

Wu Q, Andolfatto P, Haunerland NH. 2001. Cloning and sequence of the gene encoding the muscle fatty acid binding protein from the desert locust, *Schistocerca gregaria*. *Insect Biochem Mol Biol*. 31:553–562.

Yue L, Karr TL, Nathan DF, Swift H, Srinivasan S, Lindquist S. 1999. Genetic analysis of viable *Hsp90* alleles reveals a critical role in *Drosophila* spermatogenesis. *Genetics* 151:1065–1079.

Zhang DX, Yan LN, Ji YJ, Hewitt GM, Huang ZS. 2009. Unexpected relationships of substructured populations in Chinese *Locusta migratoria*. *BMC Evol Biol*. 9:144.

Zhang J, Wyatt GR. 1996. Cloning and upstream sequence of a juvenile hormone-regulated gene from the migratory locust. *Gene* 175:193–197.