BMC Genomics

RESEARCH

Open Access

Comparative genomics uncovers evolutionary drivers of locust migratory adaptation

Sujuan Dong[1], Xinghua Li[1], Qi Liu[2], Tao Zhu[3], Aiwei Tian[1], Nuo Chen[1], Xiongbing Tu[4*] and Liping Ban[1*]

Abstract

Background Locust migration is one of the main causes of locust plagues. While existing research has highlighted the adaptive migratory capabilities of locusts, the evolutionary patterns of their migration remain elusive. This study aims to explore these evolutionary patterns of locust migratory behavior at the genomic level. To achieve this, we conducted comparative genomics analysis using genomic data from 10 locust species with diverse migratory tendencies.

Results We identified 1064 genes showing signatures of positive selection in five migratory locust species using a dN/dS model. The BUSTED-PH model revealed 116 genes associated with migratory phenotypes. Gene ontology enrichment analysis indicated that these genes were predominantly related to metabolism and mitochondria-related pathways through both methods. Additionally, the evolutionary rate (RER) analysis between migratory and non-migratory locusts revealed significant divergence in energy metabolism pathways. Notably, of the genes analyzed, the *SETX* gene consistently showed evidence of positive selection across all five migratory species.

Conclusions The findings suggest that the evolution of migratory behavior is associated with increased selective pressure on metabolism and mitochondria-related pathways. Hundreds of genes undergo selective changes during repetitive transitions to migratory behavior. These findings enhance our understanding of the genetic and phenotypic relationships underlying different locust migratory behaviors, providing important data for understanding the biological mechanisms behind locust outbreaks.

Keywords Locust migration, Evolution, Positive selection, Metabolism, Mitochondria-related pathways

*Correspondence:
Xiongbing Tu
txb1208@163.com
Liping Ban
liping_ban@163.com
[1] Department of Grassland Resources and Ecology, College of Grassland Science and Technology, China Agricultural University, Beijing 100193, China
[2] Rice Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, Guangdong Province 510640, China
[3] State Key Laboratory of Animal Nutrition, Department of Animal Genetics and Breeding, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China
[4] Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100193, China

Background

Grasshoppers, being predominant herbivores, have diversified and adapted to various habitats, including grasslands, deserts, semi-aquatic zones, alpine regions, and tropical forests, demonstrating a broad spectrum of morphological, ecological, and behavioral variations [1]. They play a crucial role in grassland ecosystems by influencing nutrient cycling [2]. According to a biogeographical analysis, grasshoppers originated in South America. Throughout the diversification of the Acrididae family, there have been numerous instances of colonization and recolonization between the New World and the Old World [2–4]. The subfamily Cyrtacanthacridinae, Gomphocerinae, Melanoplinae, and Oedipodinae exhibit a widespread distribution across the globe [5]. The grasshopper family Acrididae includes over six thousand species distributed globally [1, 4, 6], there are significant differences in the migratory abilities of different grasshopper species. Some species are flightless or non-migratory, while others are migratory. Migratory species such as *Schistocerca gregaria*, *Schistocerca cancellata* [7], *Schistocerca piceifrons* [8], *Locusta migratoria* [9], *Calliptamus italicus* [10], and *Gomphocerus sibiricus* [11], can form swarms and migrate over thousands of kilometers, causing extensive damage to plants and crops along their migratory route. Non-migratory species, such as *Stenobothrus lineatus* [12], *Aidemona azteca, Xenocatantops brachycerus* (catantopinae), *Schistocerca nitens* and *Schistocerca americana* [13, 14], move within their native habitat and disperse to adjacent suitable locations through walking or jumping. The *Stenacris vitreipennis* species and the genus *Conophyma zubovsky* are flightless [15], characterized by either being wingless or having short wings.

Locusts are a special type of grasshopper (Orthoptera: Acrididae) [1]. Approximately 20 species recognized as true locusts [16]. The definition of locust is based on two principal criteria: (1) they form dense groups and migrate, and (2) they exhibit polymorphism, with individuals in solitary conditions differing in many characteristics from those in groups [16–18]. The *L. migratoria*, *S. gregaria, S. piceifrons, S. cancellate, Nomadacris septemfasciata, Locustana pardalina, C. italicus* [19] are called locusts due to their striking changes in color, behavior, morphology, biochemistry, and life history traits in response to change in local population density.

The genus *Schistocerca* (Orthoptera: Acrididae: Cyrtacanthacridinae) includes 50 species. The majority of *Schistocerca* species do not swarm, only the desert locust (*S. gregaria*), the Central American locust (*S. piceifrons*), and the South American locust (*S. cancellata*) are considered to be locusts [19]. Only the *S. gregaria* is found in the Old World and is the earliest diverging lineage within the genus *Schistocerca*, whereas all the remaining species are distributed in the New World [1, 20]. Through ancestral character reconstruction of reaction norms, *S. gregaria*, *S. piceifrons*, and *S. cancellate* did not form a monophyletic group [20], suggesting that swarming locusts have evolved multiple times within *Schistocerca*. In the ancestral *Schistocerca* genus, the migratory behavioral plasticity of the locusts within the genus has been lost and regained at least twice [1, 20].

Locusts, as a subset of grasshoppers (Orthoptera: Acrididae), have earned a notorious reputation due to their capacity to cause extensive damage to vegetation during their migratory phases [21]. Many studies have been performed to explore control methods for locust plagues based on the flight characteristics and swarms in migratory locusts. Recent research has shown that locust aggregation is influenced by the pheromone 4-vinylanisole (4VA), which is secreted when the locust density reaches four individuals per (10 cm × 10 cm × 10 cm) volume, promoting further aggregation [22, 23]. Compared with other locust species, the gregarious locust *L. migratoria* has higher levels of hyperlipidemia and lipid storage. Additionally, metabolome—transcriptome analysis has revealed that reactive oxygen species (ROS) generation is related to the flight muscle's catabolic capacity in locusts. Solitary locusts have a higher catabolic capacity and produce more ROS during high-velocity flights, while gregarious locusts have a lower catabolic capacity and generate less ROS during long—distance flights [21]. In the Tibetan Plateau, migratory locusts adapt to the high-altitude hypoxic environment by reducing their body size and with the help of the gene *PTPN1*, which participates in carbon and energy metabolism [24].

Research on the evolutionary relationships of migratory locusts has revealed that migratory locusts are distributed among different species [1, 4, 25]. For example, *L. migratoria* belongs to the subfamily Oedipodinae [25], whereas desert locusts (*S. gregaria* [25], *S. cancellate* and *S. piceifron*) belong to the subfamily Cyrtacanthacridinae [1, 26]. Convergent evolution refers to the phenomenon in which different species independently evolve similar traits [27, 28]. The independent multiple evolutions of migratory traits within the *Schistocerca* genus imply the convergent evolution of migratory traits in the Acrididae family. However, the genetic basis and molecular mechanisms underlying migratory traits in locusts are not well understood.

With advancements in genomic technologies, the genomes of various locust lineages have been sequenced. This enables the identification of key regulatory genes related to locust migration traits by means of selection pressure analysis and evolutionary rate assessment. This research aims to explore the genetic basis of migratory

behavior more comprehensively, providing a deeper understanding of this complex evolutionary trait and its implications for pest management and ecological dynamics.

Methods

Data collection

The migratory capabilities of locusts were sourced from encyclopedias and published research. The species information is presented in Table 1, which includes 5 migratory locusts and 5 non-migratory locusts. The migratory species are *C. italicus*, *L. migratoria*, *S. cancellata*, *S. gregaria* and *S. piceifrons*, while the non-migratory species are *A. azteca*, *S. americana*, *S. nitens*, *S. lineatus* and *X. brachycerus*. There are literary records regarding their migratory abilities. The *C. italicus* locust has two forms: hopper bands and swarms. Both are capable of migration and can cover varying distances. Hopper bands typically move up to several hundred meters, while swarms can travel distances up to 100—200 km [29–32]. Regarding *L. migratoria*, the literature documents that it can fly continuously for about 12 h and spread across 60 countries [33, 34]. For *S. cancellate*, the literature records that it can travel several hundred kilometers each day and consume a large amount of vegetation [7]. For *S. gregaria* and *S.piceifrons* multiple literary sources record that they are capable of migrating over long distances [8, 9, 35]. For non-migratory locusts, the nymphs of *A. azteca* have underdeveloped wings, while adults are fully winged and are excellent jumpers [36]. *S. americana* can fly, but it seldom forms swarms [14, 37]. *S. nitens* is usually solitary and non-migratory [38]. *S. lineatus* is a non-migratory species, wing monomorphic grasshopper [12]. In terms of behavior and physiological structure, *X. brachycerus* can fly but do not possess migratory capabilities [39, 40]. The species with sequencing data were chosen based on their migratory traits. In addition, *Gryllus bimaculatus* was used as an outgroup. When selecting genetic data of species for subsequent analysis, considering all species analyzed in the study, the order of data utilization priority is as follows: 1) complete genome assembly, 2) transcriptome assembly (transcriptome shotgun assembly, TSA), 3) transcriptome sequencing.

Data preprocessing

Six species with genome assemblies, four species with transcriptome shotgun assemblies, and one species with transcriptome sequencing data remained. For the transcriptome sequencing data, FASTP v0.23.2 [43] was used to trim the low-quality reads, and subsequently, Trinity v2.15.1 [44] with default parameters was used to assemble the transcripts. For all the transcript assemblies, BUSCO v.5.3.2 [45, 46] was used to evaluate the quality and completeness of assembled transcriptomes. CD-HIT v4.6.8 [47, 48] with a similarity threshold of 0.9 was applied to remove the redundant transcripts for each species. The N50 software (https://github.com/fg6/n50) was employed to assess the N50 values of the transcriptomes. Open reading frame (ORF) identification was performed by using TransDecoder v5.7.1, Hmmscan v3.5, and BLASTX was utilized to predict the protein sequences.

Homologous gene identification

Proteinortho v6 was employed for the identification of homologous genes. Gene pairs with blast *P* values less than 1e-10 and minimum coverage exceeding 70% were identified as homologous genes [49]. Subsequently, homologous genes present in at least seven species were selected for further investigation. PRANK [50] with default parameters was utilized for sequence alignment. After alignment, PAL2NAL v14 was used to

Table 1 Genetic Information of Locusts in Different Migratory States

Species name	Abbreviation	Subfamily	Accession number	Migrate	References
Aidemona azteca	*A. azteca*	*Melanoplinae*	GIGD01	N	[36]
Calliptamus italicus	*C. italicus*	*Calliptaminae*	SRR6113310; SRR6113309	Y	[10]
Locusta migratoria	*L. migratoria*	*Oedipodinae*	GCA_026315105.1	Y	[33, 41]
Schistocerca americana	*S. americana*	*Cyrtacanthacridinae*	GIOT01	N	[14]
Schistocerca cancellata	*S. cancellata*	*Cyrtacanthacridinae*	GCA_023864275.2	Y	[7]
Schistocerca gregaria	*S. gregaria*	*Cyrtacanthacridinae*	GCA_023897955.2	Y	[9]
Schistocerca nitens	*S. nitens*	*Cyrtacanthacridinae*	GCA_023898315.2	N	[38]
Schistocerca piceifrons	*S. piceifrons*	*Cyrtacanthacridinae*	GCA_021461385.2	Y	[8]
Stenobothrus lineatus	*S. lineatus*	*Gomphocerinae*	GAUZ02	N	[12]
Xenocatantops brachycerus	*X. brachycerus*	*Catantopinae*	OFSG01	N	[40]
Gryllus bimaculatus	*G. bimaculatus*	*Gryllinae*	GCA_017312745.1	–	[42]

convert the protein alignments into codon alignments based on the corresponding nucleotide sequences [51].

Phylogenetic analyses and divergence time estimation

To reconstruct the evolutionary relationships among locusts, a total of 7861 coding sequences (CDS) of all species were concatenated using catfasta2phyml v.1.2.0. Subsequently, TrimAI v1.5.0 was employed to trim the sequence with a coverage lower than 70%. The maximum-likelihood (ML) phylogenetic tree was constructed with IQ-TREE v 2.2.2.6 [50, 52], and the support rate for the nodes was estimated based on 1500 ultrafast bootstrap replicates. To estimate the divergence time of locusts, we assumed that the phylogenetic tree obtained by IQ-TREE represented the true evolutionary relationship, and the MCMCTREE program of PAML v 4.10.7 [53] was utilized to assess the divergence time. The priori divergence times of *C. italicus*, *G. bimaculatus* and *X. brachycerus* were retrieved from the Timetree [54] database.

Positive selection analyses associated with migratory evolution

Environmental changes play a crucial role in animal trait changes. In the evolution of locust migration traits, selection pressure impacts reproduction, theoretically, related genes should have a high evolutionary rate. Therefore, we attempt to identify migration-related genes by searching for positive selection in protein-coding sequences. Firstly, we generated 5 CDS datasets based on five migratory locust species. Each of the datasets comprised only one migratory locust and five non-migratory locusts. For each dataset, we initially concatenated all the sequences using catfasta2phyml. Subsequently, we constructed a phylogenetic tree with IQ-TREE software. Then, we calculated the nonsynonymous (dN) to synonymous substitutions (dS) per site (dN/dS) values for each gene through the Codeml program of PAML package [55]. During this process, we designated the same five non-migratory species as the foreground branches instead of the migratory ones. Furthermore, to enhance the robustness of our results, we employed the branch-site unrestricted statistical test of episodic diversification and association with phenotype (BUSTED-PH) model, which integrates four models and hypotheses to investigate the correlation between selection pressure and the phenotypic manifestations associated with of coding sequences, HyPhy [56] was utilized to perform the above analysis and the specific detection steps are detailed at (https://github.com/veg/hyphy-analyses/tree/master/BUSTED-PH). During the detection process, we utilized the same five datasets, and migratory locusts were labeled as foreground branches in the phylogenetic tree.

Analysis of relative evolutionary rates associated with migration

The evolutionary rates of homologous proteins across different species can be quantified using the synonymous substitution rate and the nonsynonymous substitution rate [57], and the relative evolutionary rate can be used to explore the relationship between genes and convergent traits. In this study, we used the relative evolutionary rate (RER) [58] model of Hyphy software to investigate the relationship between migratory traits and genes in locusts. The specific detection method can be found at (https://github.com/veg/hyphy-analyses/tree/master/RER). Generally, this model requires an aligned protein sequence, an evolutionary tree, and a list of species with convergent traits designated as foreground branches. The evolutionary tree used in this analysis is constructed based on concatenated gene sequences. To ensure consistency between the evolutionary tree and the input protein sequence, we trimmed the evolutionary tree using NWKIT's [59] intersection function to make sure that the species in the protein sequence correspond with those in the evolutionary tree.

Gene ontology (GO) term enrichment analysis

For each homologous gene, the longest homologous protein was extracted. Subsequently, the sequence was uploaded to the eggNOG-mapper database [60, 61] for gene function annotation. Then, gene function enrichment analysis was performed via the clusterProfiler software package in R [62].

Results

Locust transcriptome assembly

In this study, no available TSA data for *C. italicus* could be found. Hence, public transcriptome sequences were utilized to assemble the transcripts. After assembly process, the total length of the contigs amounted to 783,485,549 bp, comprising a total of 1,292,693 contigs. The maximum and minimum contig lengths were 63,669 bp and 266 bp respectively. The N50 value was 777 bp, and the GC content was 40.93%. The transcriptome assembly of four other species were also evaluated (Table 2). Among them, *X. brachycerus* presented the highest transcriptome assembly quality, with an N50 value of 1799 bp. The BUSCO evaluation indicated that *C. italicus* had the highest transcriptome assembly completeness, reaching up to 97.6%, while the completeness of the other four species exceeded 71%. Using

Table 2 Summary of Transcriptome Assembly Results for different grasshopper species

Sample	Full length(bp)	Contigs	GC (%)	N50	Max contig	Min contig	Complete (%)
A. azteca	85,250,201	138,479	40.29	934	10,117	241	71.6
C. italicus	783,485,549	1,292,693	40.93	777	63,669	266	97.6
S. americana	160,912,811	305,724	41.16	622	43,436	228	94.3
S. lineatus	137,719,092	387,115	40.71	281	22,782	200	79.1
X. brachycerus	41,642,132	43,187	42.7	1799	19,641	207	85.5

Proteinortho, a total of 35,312 homologous genes were identified. Among these genes, 7861 homologous genes were covered by at least 7 species. Of these, 6,874 genes were successfully annotated to known genes (Additional file S1).

Phylogenomics

Based on 7861 homologous genes, the phylogenetic relationships of 10 grasshopper species were constructed using IQ-TREE, which strongly supported all nodes (bootstrap values ≥ 98). Among the phylogenetic trees, *A. azteca* (subfamily of Melanoplinae) was the most distant species from the other nine grasshoppers. The species *S. americana*, *S. cancellata*, *S.gregaria*, and *S. nitens* were clustered into the Cyrtacanthacridinae subfamily clade, having a close phylogenetic relationship with the *X. brachycerus* of the Catantopinae subfamily. Within this branch, there were three species of locusts with the ability to migrate. The remaining grasshoppers, representing three different genera, were clustered on a single large branch, namely, *C. italicus* (the Calliptaminae subfamily), *S. lineatus* (the Gomphocerinae subfamily) and *L. migratoria* (the Oedipodinae subfamily). Within this branch, two locust species were capable of migration (Fig. 1).

To estimate the divergence time for each node, the following divergence times were obtained from Timetree: between *G. bimaculatus* and other locusts, it was 241.3–339.4 million years ago (MYA); between *X. brachycerus* to the rest locusts, was 62.0–66.9 MYA. Then, MCMCTREE was used to infer the divergence time of all species. The results indicated that the divergence time between *G. bimaculatus* and grasshoppers was approximately 347 MYA. Subsequently, *A.zteca* began to differentiate as a new species about 282 million years ago, between 66 and 20 million years ago, the remaining locusts began to rapidly diverge (Fig. 1).

Genome-wide patterns of accelerated molecular evolution converge in migratory locusts

In this study, we designated migratory locusts as foreground branches and non-migratory locusts as background branches with the aim of calculating the dN/dS values of all genes in different migratory locusts. The significance of the dN/dS values between background and foreground branches was subsequently evaluated through the chi-square test, and genes whose dN/dS values for migratory locusts were greater than two and whose *P* values were less than 0.05 were selected as candidate positive selection genes. Out of the 6874 homologous genes, we identified a total of 1064 genes that had undergone positive selection in at least on lineage. Notably, when considering the species *L. migratoria*, *S. piceifrons*, *S. cancellate*, *S. gregaria*, and *C. italicus*, the number of genes under positive selection were 325, 233, 229, 213, and 211, respectively (Fig. 2). Among these species, *L. migratoria* and *C. italicus* were found to share 29 positively selected genes. In addition, three migratory *Schistocerca* genus species were observed to share a total of 6 genes. Of these 6 genes, five were annotated as *Aubergine* (*AUB*), *UDP-glucuronosyltransferase gene* (*UDPGT*), *Histidine phosphatase superfamily (branch 2)* (*HIS_PHOS_2*), *Small nuclear ribonucleoprotein U5 subunit 200* (*SNRNNP200*) and *Senataxin* (*SETX*). Specific genes play a crucial role in the regulation of diverse biological processes and molecular functions, such as gene expression, cellular development, protein function, and morphogenesis. The regulation of these processes and functions is crucial for the normal development and maintenance of physiological functions in organisms (Table 3). The *actin beta* (*ACTB*) gene was detected in four migratory species. Moreover, the *SETX* gene was found to be shared among the five migratory locust species (Additional file S2).

To explore the functional trends of positively selected genes in locusts, we performed gene function enrichment analysis. For the 1064 positively selected genes (*P* values < 0.05) where migratory locusts were set as the foreground branch and non-migratory locusts as the background branch, 133 GO terms were annotated (*P* values < 0.05) (Additional file S3), In the 20 most significant GO terms, 18 were directly related to the metabolic function (Fig. 3A). For the 1889 positively selected genes (*P* values < 0.05) where non-migratory locusts were set as the foreground branch and migratory locusts as the background branch, 413 GO terms were annotated (*P*

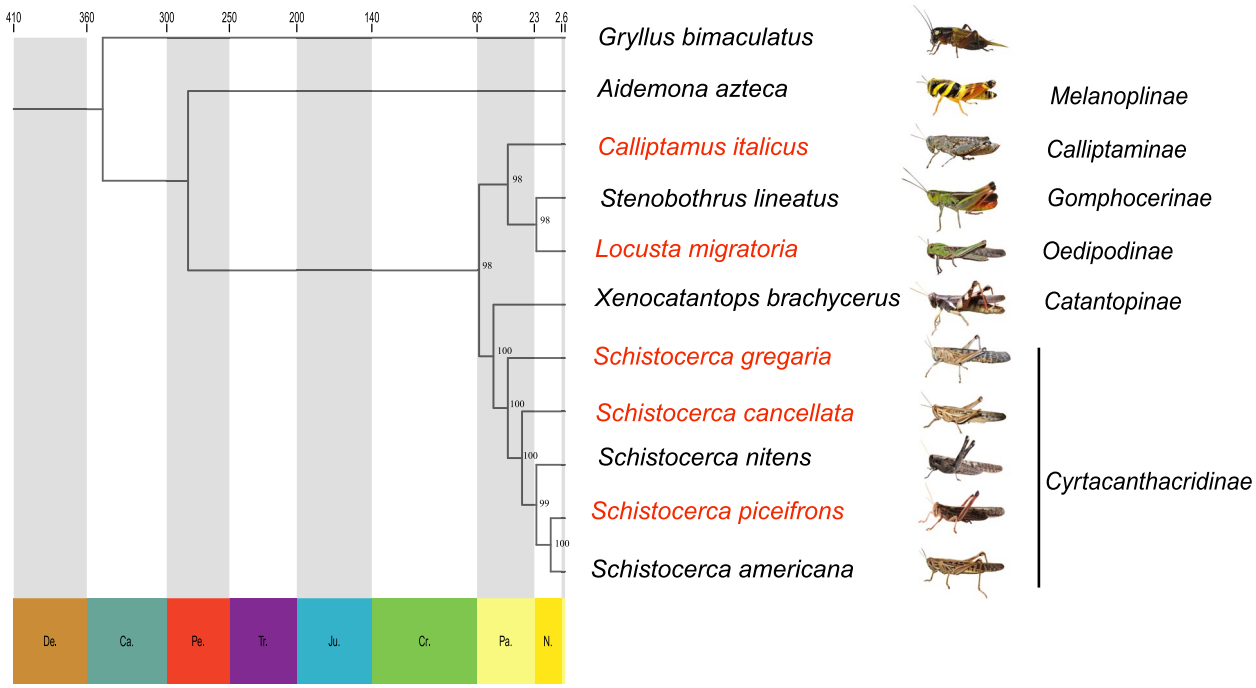

Fig. 1 Phylogenetic trees and estimated divergence times of the 10 grasshopper species. Red species represent locusts with migratory ability. The left side shows the divergence time of the locusts. The right side displays the phylogenomic information of locusts. The numbers at the branch nodes of the phylogenetic tree represent the support rates, which reflect the reliability of the branches

values < 0.05) (Additional file S4), we took the top 20 GO terms. Among them, merely 8 were associated with metabolic functions (Fig. 3B).

Through dN/dS analysis, we have identified 124 genes that were detected to be under positive selection in at least two migratory locusts. At the GO term level of these genes, there were 117 GO terms associated with locust migratory behavior (*P* value < 0.05) (Additional file S5), and six of them were related primarily to mitochondrial components, such as the mitochondrial inner membrane (GO:0005743) (Fig. 3C). Among the genes associated with this GO term (GO:0005743), we further identified 10 out of 124 genes that presented strong evidence of selection pressure in migratory species: *ABCB10*, *FOXRED1*, *HIGD1A*, *MRPL13*, *MRPS22*, *MRPS9*, *MTG1*, *NDUFB5*, *NDUFC2*, *UQCRFS1* (Additional file S6). The genes *NDUFC2*, *FOXRED1*, *NDUFB5*, *UQCRFS1*, and *ABCB10* are related to the inner membrane of mitochondria. *MRPS22*, *MRPS9*, and *MRPL13* are components of mitochondrial ribosomes. *HIGD1A* is involved in transmembrane transport. *MTG1* plays a role in regulating the assembly and translational activity of mitochondrial ribosomes and is related to ribosomes, mainly located in the mitochondrial matrix. Some of these proteins are located on the inner membrane of mitochondria and participate in processes related to the respiratory chain and transmembrane transport, while others located in the

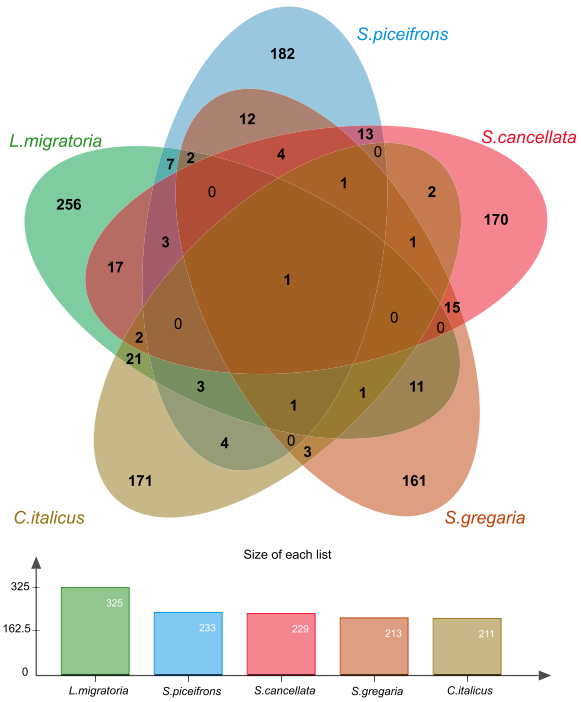

Fig. 2 Genome-wide patterns of convergent changes (dN/dS) linked to migratory evolution. In the Venn diagram, the numbers within the overlapping regions represent the quantities of positive-selected genes shared among different migratory species, while the numbers in the non-overlapping regions represent the quantities of positive-selected genes unique to each species. The bar chart represents the total numbers of positive-selected genes of different migratory species identified by the dN/dS method

Table 3 The functions of 6 positively selected genes of locusts identified by dN/dS method

Gene full name	Symbol	Species	Gene description
Aubergine	*AUB*	*S.gregaria*, *S.piceifrons*, *S.cancellata*	Plays a crucial role in germline formation, mRNA localization and translational regulation, as well as gene silencing
UDP-glucuronosyltransferase gene	*UDPGT*	*S.gregaria*, *S.piceifrons*, *S.cancellata*	Encodes enzymes that catalyze the glucuronidation of a wide range of endogenous and exogenous substances, playing a crucial role in detoxification and metabolism
Small nuclear ribonucleoprotein U5 subunit 200	*SNRNP200*	*S.gregaria, S.piceifrons, S.cancellata*	Involved in mRNA splicing, essential for the formation of the spliceosome complex that removes introns from pre-mRNA
Histidine phosphatase superfamily (branch 2)	*HIS_PHOS_2*	*S.gregaria, S.piceifrons, S.cancellata, C.italicus*	Encodes enzymes that catalyze the dephosphorylation of various substrates, playing a crucial role in cellular signaling and regulation
Senataxin	*SETX*	*S.gregaria, S.piceifrons, S.cancellata, C.italicus, L.migratoria*	*Encodes the senataxin protein, which is crucial for DNA repair, transcription, and maintaining genomic stability*
Actin beta	*ACTB*	*S.gregaria, S.piceifrons, C.italicus, L.migratoria*	Encodes one of the six actin proteins. These highly conserved actins are involved in cell motility, structure, integrity, and signaling

mitochondrial matrix and participate in ribosome assembly and protein translation processes, providing energy for locust migration (Fig. 4). The selection signatures in these 10 mitochondrial genes suggest their role in locust migration. 9 out of 117 GO terms were associated with T-cell functions (Additional file S5), including the regulation of T-cell differentiation (GO:0045580) (Fig. 3C). The gene *PNP* was associated with 8 out of 9 GO terms related to T-cell functions. It was present in *S. piceifrons*, *C. italicus*, and *L. migratoria*. *PNP* encodes an enzyme that is responsible for the reversible phosphorolysis of purine nucleosides [63, 64]. This indicates that *PNP* may play a key role in the migratory ability evolution of locust.

Apart from using the dN/dS method, we employed a stricter model, BUSTED-PH, to assess selection pressures related to migratory behavior. This model evaluated whether the sequence evolution rate of genes was correlated with the presence of migratory traits in these five locust species. The results indicated that 116 genes exhibited phenotypic associations according to the BUSTED-PH model. Furthermore, the gene function enrichment analysis (Additional file S7) revealed 160 GO terms that were significantly linked to locust migratory behavior (P value < 0.05). Among them, 15 terms were associated with mitochondria-related pathways, such as the mitochondrial inner membrane (GO:0005743), the mitochondrial protein-containing complex (GO:0098798), the mitochondrial translation (GO:0032543), the mitochondrial ribosome (GO:0005761), the mitochondrial electron transport, and the conversion of NADH to ubiquinone (GO:0006120) (Fig. 3D). The consistency between the two models strongly suggests that there is positive selection pressure on mitochondrial energy metabolism genes in migratory locusts. Combining the results of the two analyses, a total of 61 migration-related genes were identified via both the dN/dS and BUSTED-PH methods (Fig. 5). Among these genes, 20 were specific to *C. italicus*, 15 to *L. migratoria*, 8 to *S. piceifrons*, 5 to *S. gregaria*, and 4 to *S. cancellata* species. Additionally, three migration-related genes, namely *NDUFC2*, *SND1*, and *Br–c*, were shared between *L. migratoria* and *S. gregaria*, and the *GlcAT-P* and *Vha16* genes were found to be shared between *C. italicus* and *L. migratoria*. Furthermore, *CES5A* was identified as a migration-related gene shared among *C. italicus*, *S. cancellata*, and *S. gregaria* (Additional file S8).

To further identify potential convergent evolutionary genes in migratory locusts, we calculated the evolutionary rate between migratory locusts and non-migratory locusts using the RER model. Interestingly, 392 genes exhibited a significantly higher evolutionary rate (P value < 0.05) in migratory locusts than in non-migratory locusts (Additional file S9). By integrating the RER, dN/dS, and BUSTED methods, we identified a total of 1405 genes associated with migratory traits. Specifically, 14 genes were identified in both RER and BUSTED, 100 genes in both RER and dN/dS, and 61 genes in both BUSTED and dN/dS. Additionally, 8 genes were identified in all three methods (Fig. 6). To further explore the functional trends of genes involved in migratory behavior evolution, we focused on the subset of genes with consistent behavior between the RER and dN/dS

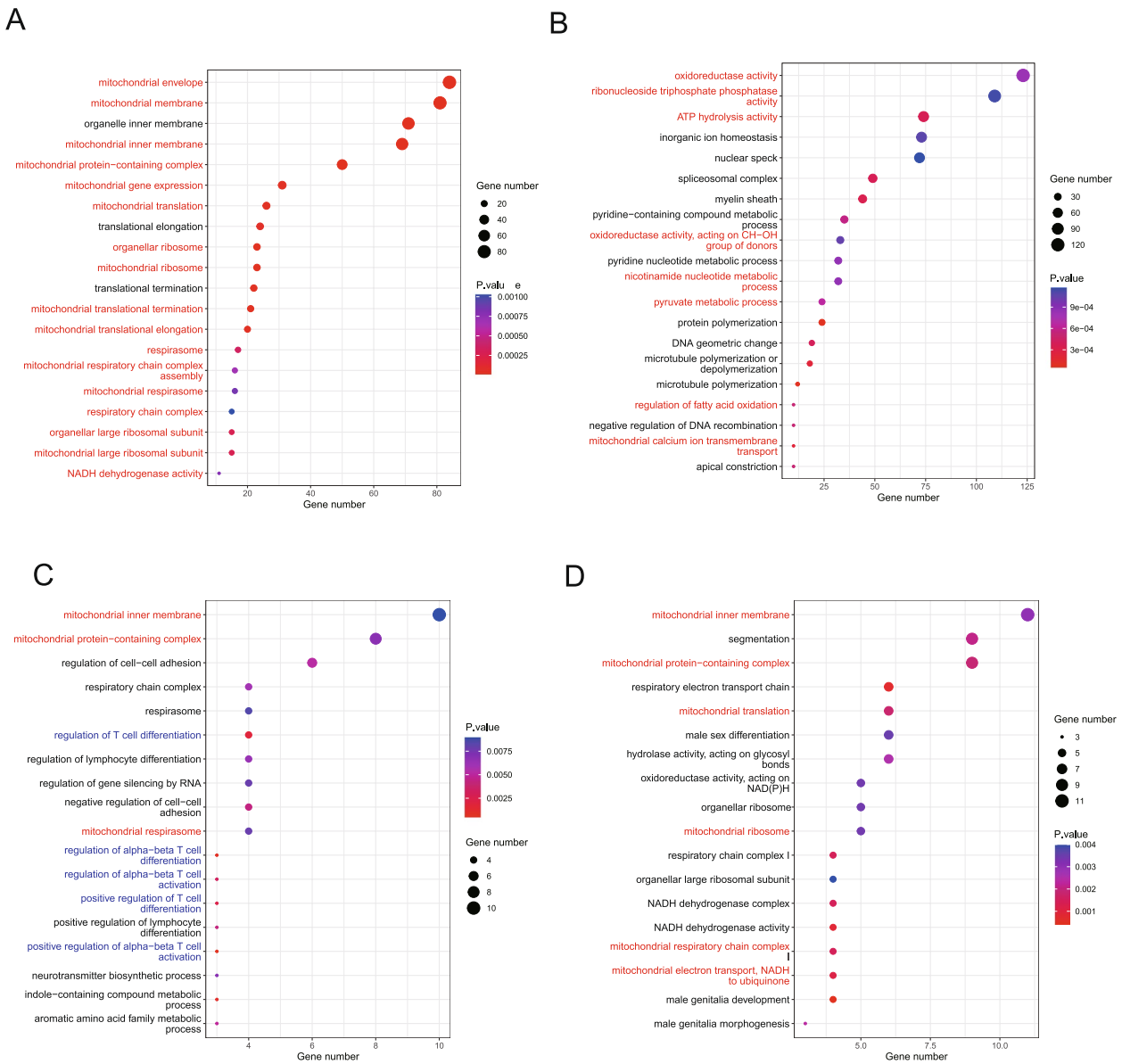

Fig. 3 The GO enrichment analysis of positively selected genes. **A**, **B** and **C** display the GO enrichments analysis of positively selected genes obtained through the dN/dS method. **A** The GO enrichment of positively selected genes in five migratory locust species. **B** The GO enrichment of positively selected genes in five non-migratory locust species. In A and B, the red GO terms are correlated with energy metabolic pathways. **C** The GO enrichment of positive selection genes in at least two migratory locusts. **D** The GO enrichment of positive selection in migratory branches that by the BUSTED-PH method. In **C** and **D**, red represents pathways associated with mitochondria, while blue represents pathways associated with T-cells

models. This yielded 100 genes with a higher evolutionary rate, among which only six genes were present in at least three of the migratory species, including genes *AUB*, *SNRNP200*, *SETX*. Among the significantly associated genes, selection signal of *SETX* gene was consistently present in all five migratory species. In addition, the *LOC126204421 gene*, involved in oxidoreductase activity (GO:0016491), flavin adenine dinucleotide binding (GO:0050660), organic hydroxy compound metabolic process (GO:1,901,615), and pyridine-containing compound metabolic process (GO:0072524), was specifically identified in *S.gregaria*, *C. italicus*, and *L. migratoria* (Additional file S10).

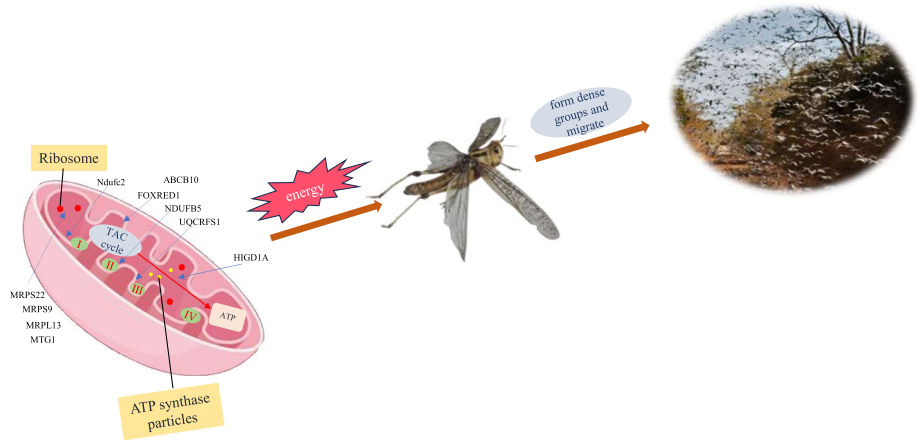

Fig. 4 The 10 positive selection genes within the mitochondrial inner membrane (GO:0005743) GO term

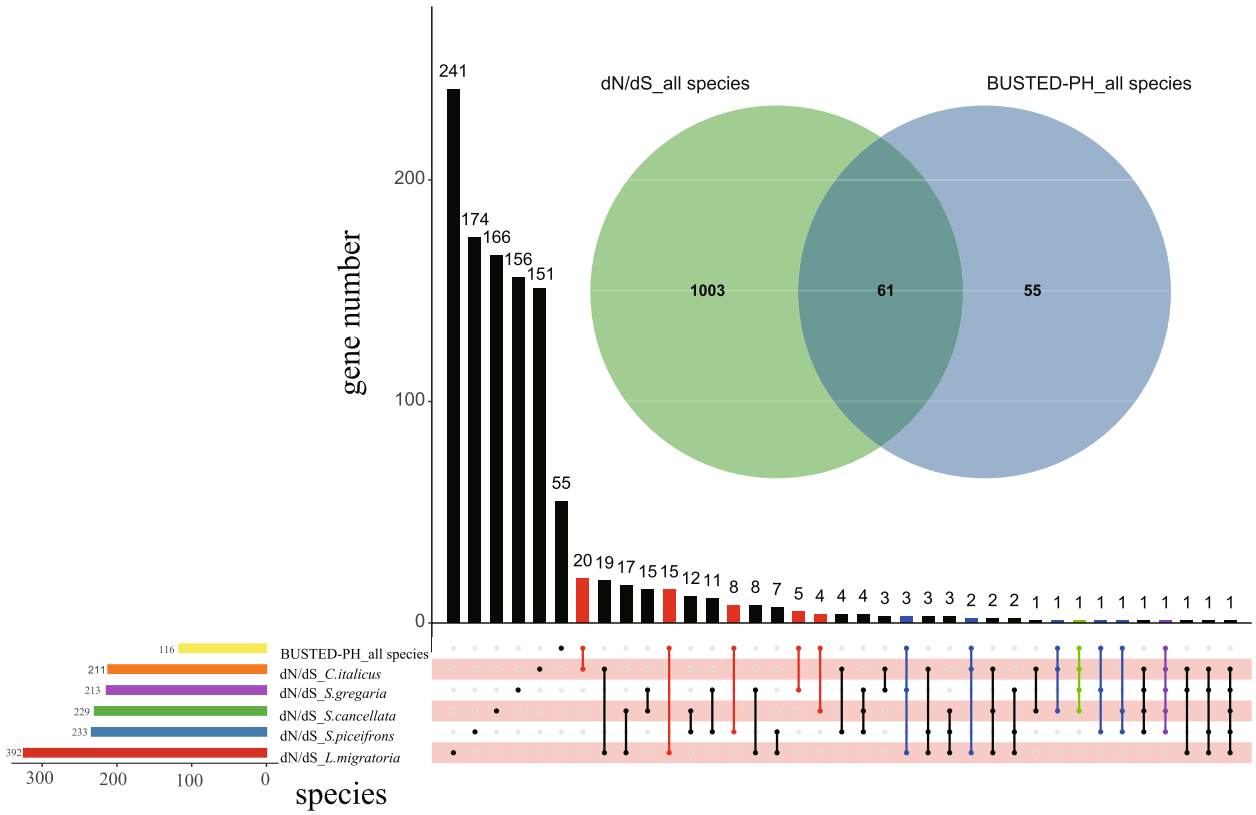

Fig. 5 Joint identification of positive genes related to migration. Venn diagrams represent the number of positively selected genes identified by the dN/dS method and the BUSTED—PH method. The bars display the number of positive-selection genes identified by different methods. The combination of vertical lines and solid dots represents a set of genes related to migratory-related positive selection genes obtained by dN/dS analysis of different migratory locusts and genes corresponding to locusts and migratory traits found by the BUSTED—PH method

Discussion

In our study, we used a comparative genomic approach involving 10 grasshopper species with different migratory capacities to investigate the genomic changes associated with the convergent evolution of migration. These species effectively cover nearly all known locust ecological niches related to flight, which have important implications for analyzing complex behaviors such as migration.

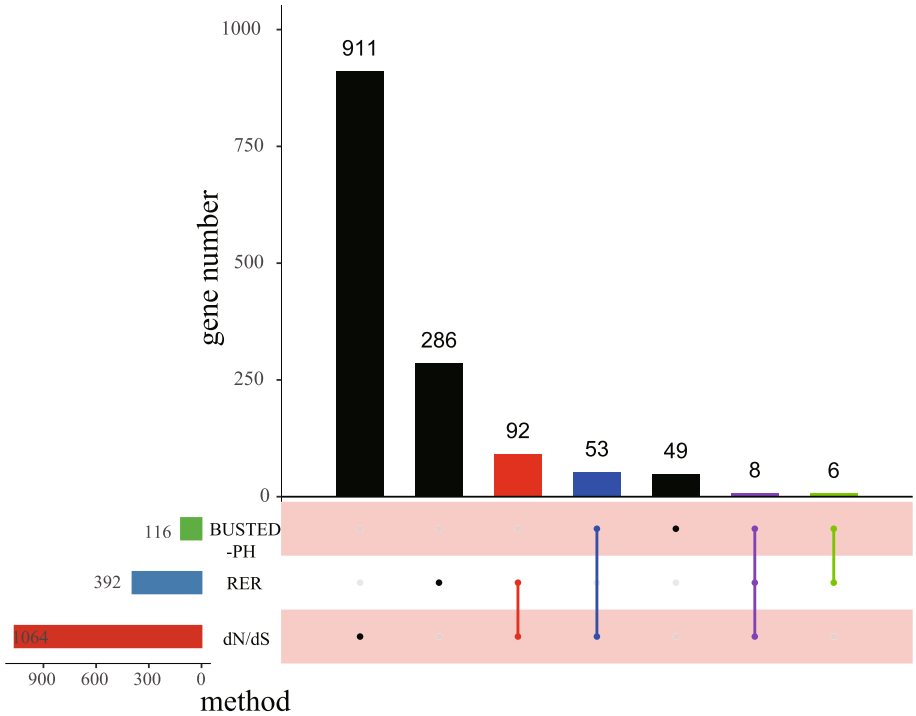

Fig. 6 Convergent evolutionary analysis of genes associated with locust migration. The bars display the number of genes identified by different methos related to locust migratory traits. The combination of vertical lines and solid dots represents sets found by different methods

Identifying positive selection is challenging because it occurs sporadically on a few specific amino acid sites, and its signal may be obscured by negative selection. In our research, the identification of genes involved in migration evolution mainly depend on selection pressure. The principal computational models employed for estimating these pressures are PAML's codeml and HyPhy's BUSTED. The branch-site model in PAML and BUSTED-PH for positive selection has a low false-positive rate and offers greater statistical power than branch-based tests [65–67]. BUSTED provides a genome-wide (non-site-specific) test for positive selection by checking whether a gene has undergone positive selection at least at one site in at least one branch. Our analysis, which uses dN/dS and HyPhy BUSTED-PH to assess selective pressure across the genome, revealed that specific genes are associated with metabolism and mitochondria-related functions. These molecular adaptations probably reflect locust specialization in migratory behavior, and several candidate genes stand out in this respect.

Biologists have long been interested in convergent evolution, like mammalian vocal learning [68], bird and bat echolocation [69], and the adaptation of marine mammals to extreme marine environments [70]. To survive in new environments, animals may completely change their habits, and migration is a common phenomenon among animals [71–73]. Although the functional nature of these convergent specializations is often obvious, the genetic basis underlying specific examples of convergent evolution is far less clear [74].

In our research, we identified numerous genes associated with migratory traits. However, a larger number of significant genes were identified primarily on basis of strong evidence within one of the migratory clades, while the evidence was weaker in the other four. Eventually, only 6 out of 392 of the significantly associated genes presented robust evidence of differential rates of evolution in at least three of the migratory clades. Although hundreds of important genes have been identified, only the *SETX* gene was supported by strong evidence from all migratory lineages. The *SETX* gene encodes a protein that has RNA helicase activity and is encoded by a domain at the C-terminal end of the protein. The protein encoded by this gene contains a DNA/RNA helicase domain at its C-terminal end, which suggests that it may be involved in both DNA and RNA processing. The most relevant gene, *SETX*, represents a particularly promising candidate as it had been previously associated with muscle development [75]. This gene is homologous to yeast *SEN1* and exhibits DNA/RNA helicase activity; the analysis of R-loop structures revealed that it is related to transcriptional regulation and the DNA damage response [76, 77].

Many neurodegenerative diseases are associated with the formation and resolution of R-loops, and the *SETX* gene is closely linked to the autophagy pathway [78]. It uniquely contributes to normal muscle function, and depletion of *SETX* inhibits the progression of autophagy, resulting in mitochondrial defects. Dominant mutations in *SETX* lead to a juvenile form of amyotrophic lateral sclerosis (ALS) known as ALS4, whereas recessive mutations result in a condition called ataxia with oculomotor apraxia type 2 (AOA2) [75]. Previous studies have shown that insect flight involves neuroendocrine control and rhythmic muscle contractions [79]. The *ACTB* gene has shown strong evidence of convergent selective pressure in *S.gregaria*, *S. piceifrons*, *C. italicus* and *L. migratoria*. *ACTB* encodes one of five different actin proteins and is an abundant and highly conserved cytoskeletal protein that plays critical roles in cell functions such as migration, division, and gene expression because it has the ability to form filaments that can be rapidly assembled and disassembled in response to the needs of cells [80]. It has long been recognized as a common housekeeping gene and reference protein [81].

In our study, the convergent evolution observed in migratory locusts was correlated with mitochondrial adaptations. Mitochondria are the primary organelles responsible for generating ATP via oxidative phosphorylation, which reflects their evolutionary changes in oxidative phosphorylation capacity and adaptation to the cellular energy demand. This evolutionary process reflects both their capacity for oxidative phosphorylation and adaptation, which is consistent with prior research on locust flight capabilities [82]. Although various methods have been employed to detect genome-wide convergent evolution, detecting convergent amino acid substitutions remains a challenge.

Conclusions

In conclusion, through selection pressure analysis and convergent evolution analysis, we revealed that genes related to energy metabolism, mitochondrial ATP generation were under selection, indicating energy might constitute the physiological basis for locusts to adapt to the migratory environment. At the pathway level, almost all migratory locusts exhibited energy-related genes under positive selection, indicating that there was a convergent evolution phenomenon at the molecular function level in migratory locusts. However, this study did not identify specific convergent evolution sites. This finding indicates that during the locust migratory process, a common selective pressure on multiple genes, rather than a single gene, might determine locust migratory traits.

Abbreviations

BUSTED-PH	Branch-site unrestricted statistical test of episodic diversification and association with phenotype
CDS	Coding sequence
dN/dS	Nonsynonymous (dN) to synonymous substitutions (dS) per site
ORF	Open reading frame
RER	Relative evolutionary rate
TSA	Transcriptome shotgun assembly

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11376-5.

Supplementary Material 1: Additional file showing the Table S1 to Table S10. S1: The functional annotation results of homologous genes. S2: The 35 positively selected genes shared by *C.italicus*, *L.migratoria*, *S.gregaria S.piceifrons* and *S.cancellata*. S3: The gene enrichment analysis results for the 1064 positively selected genes in migratory locusts. S4: Gene enrichment analysis results for the 1889 positively selected genes in non-migratory locusts. S5: Gene enrichment analysis results for the 124 genes detected to be under positive selection in at least two migratory locusts. S6: The 10 positive selection genes within the mitochondrial inner membrane (GO:0005743) GO term. S7: The gene enrichment analysis results of 116 genes associated with migration phenotypes identified by the BUSTED-PH model. S8: Candidate genes related to migration identified by multiple analyses. S9: The list of 392 high RER gene. S10: The gene enrichment analysis results for the 392 high RER genes.

Acknowledgements

We would like to express our sincere gratitude to the High—performance Computing Platform of China Agricultural University.

Authors' contributions

L.B. conceived and designed the research. S.D. performed the experiments, analyzed and interpreted the data, and wrote the manuscript. X.L. provided L. migratoria Genomic data. Q.L. participate in wrote the manuscript. T.Z. participate in data analysis and experimental design. A.T., N.C., and X.T. participate in species data collection.

Funding

This work was supported by National Key R&D Program of China 2022YFD1400500.

Data availability

The datasets for Locusta migratoria, Schistocerca cancellate, Schistocerca gregaria, Schistocerca nitens, Schistocerca piceifrons, and Gryllus bimaculatus are available in the GenBank (https://www.ncbi.nlm.nih.gov/genbank/) under the following accession IDs: GCA_026315105.1, GCA_023864275.2, GCA_023897955.2, GCA_023898315.2, GCA_021461385.2, GCA_017312745.1. The datasets for Aidemona Azteca, Schistocerca americana, Stenobothrus lineatus, and Xenocatantops brachycerus are available in the Transcriptome Shotgun Assembly Sequence Database (https://www.ncbi.nlm.nih.gov/genbank/tsa/) under the accession IDs: GIGD01, GIOT01, GAUZ02, OFSG01. The transcriptome assembly of Calliptamus italicus was assembled based on the RNA-seq data. The raw RNA-seq data is available in Sequence Read Archive database (https://www.ncbi.nlm.nih.gov/sra) under the accession IDs: SRR6113310 and SRR6113309, and the transcriptome assembly has been deposited in figshare with the link https://doi.org/10.6084/m9.figshare.28390598.v1.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 3 September 2024 Accepted: 17 February 2025
Published online: 28 February 2025

References

1. Song H, Foquet B, Mariño-Pérez R, Woller DA. Phylogeny of locusts and grasshoppers reveals complex evolution of density-dependent phenotypic plasticity. Sci Rep-UK. 2017;7(1):6606.
2. Holland RA, Wikelski M, Wilcove DS. How and why do insects migrate? Science. 2006;313(5788):794–6.
3. Latchininsky A, Sword G, Sergeev M, Cigliano MM, Lecoq M. Locusts and Grasshoppers: Behavior, Ecology, and Biogeography. Psyche. 2011;2011: 578327.
4. Song H, Mariño-Pérez R, Woller DA, Cigliano MM. Evolution, Diversification, and Biogeography of Grasshoppers (Orthoptera: Acrididae). Insect Syst Divers. 2018;2(4):3.
5. Cigliano MM, Braun H, Eades DC, Otte D. Orthoptera Species File. In.: Species File Group; 2024.
6. Pm S, Ott SR, Niven JE. Environmental Adaptation, Phenotypic Plasticity, and Associative Learning in Insects: The Desert Locust as a Case Study. Integr Comp Biol. 2016;56(5):914–24.
7. Tumelero F, Da SV, Weymar GJ, Buske D, de Quadros RS, Gonçalves GA, de Athayde AS, Piovesan LR, Da CFI. Simulation of the Movement of a Locust Swarm in Argentina in 2020. NEOTROP ENTOMOL. 2021;50(5):716–24.
8. Ndez-Zul MIHA, Quijano-Carranza JAN, Pez RYAN, Zquez RIAV, Torres-Pacheco I, Lez RONG, Rez AECI. Dynamic Simulation Model of Central American Locust Schistocerca piceifrons (Orthoptera: Acrididae). In.; 2013.
9. Klein I, Oppelt N, Kuenzer C. Application of Remote Sensing Data for Locust Research and Management-A Review. Insects. 2021;12(3):233.
10. Sivanpillai R, Latchininsky AV, Peveling R, Pankov VI. Utility of the IRS-AWiFS Data to map the potential Italian locust (Calliptamus italicus) habitats in northeast Kazakhstan. In: 2009–1–1 2009; 2009: 9–13.
11. Wu R, Guan JY, Wu JG, Ju XF, An QH, Zheng JH. Predictions Based on Different Climate Change Scenarios: The Habitat of Typical Locust Species Is Shrinking in Kazakhstan and Xinjiang, China. Insects. 2022;13(10):942.
12. Samietz J, Köhler G. A fecundity cost of (walking) mobility in an insect. Ecol Evol. 2012;2(11):2788–93.
13. Sobel EC. The locust's use of motion parallax to measure distance. J Comp Physiol A. 1990;167(5):579–88.
14. Sword GA. To be or not to be a locust? A comparative analysis of behavioral phase change in nymphs of Schistocerca americana and S. gregaria. J Insect Physiol. 2003;49(7):709–17.
15. Zhang Y, Xu Q, Zhang D. Three new species of the genus Conophyma Zubovsky, 1898 (Orthoptera, Acridoidea, Catantopidae, Conophyminae) from Xinjiang. China Zootaxa. 2015;4039:575.
16. Pocco ME, Cigliano MM, Foquet B, Lange CE, Nieves EL, Song H. Density-Dependent Phenotypic Plasticity in the South American Locust, Schistocerca cancellata (Orthoptera: Acrididae). Annals of the Entomological Society of AmericaAnnals of the Entomological Society of America. 2019;112(5):458–72.
17. Simpson SJ, Sword GA. Locusts. Curr Biol. 2008;18(9):R364–6.
18. Foquet B, Castellanos AA, Song H. Comparative analysis of phenotypic plasticity sheds light on the evolution and molecular underpinnings of locust phase polyphenism. Sci Rep-UK. 2021;11(1):11925.
19. Ganina MD, Tyurin MV, Zhumatayeva UT, Lednev GR, Morozov SV, Kryukov VY. Comparative Analysis of Epicuticular Lipids in Locusta migratoria and Calliptamus italicus: A Possible Role in Susceptibility to Entomopathogenic Fungi. Insects. 2022;13(8):736.
20. Song H. On the origin of the desert locust Schistocerca gregaria (Forskål) (Orthoptera: Acrididae: Cyrtacanthacridinae). P Roy Soc B-Biol Sci. 2004;271(1548):1641–8.
21. Du B, Ding D, Ma C, Guo W, Kang L. Locust density shapes energy metabolism and oxidative stress resulting in divergence of flight traits. P Natl Acad Sci USA. 2022;119(1):e2115753118.
22. Guo X, Yu Q, Chen D, Wei J, Yang P, Yu J, Wang X, Kang L. 4-Vinylanisole is an aggregation pheromone in locusts. Nature. 2020;584(7822):584–8.
23. Chen D, Hou L, Wei J, Guo S, Cui W, Yang P, Kang L, Wang X. Aggregation pheromone 4-vinylanisole promotes the synchrony of sexual maturation in female locusts. Elife. 2022;11:e74581.
24. Ding D, Liu G, Hou L, Gui W, Chen B, Kang L. Genetic variation in PTPN1 contributes to metabolic adaptation to high-altitude hypoxia in Tibetan migratory locusts. Nat Commun. 2018;9(1):4991.
25. Song H, Amédégnato C, Cigliano MM, Desutter-Grandcolas L, Heads SW, Huang Y, Otte D, Whiting MF. 300 million years of diversification: elucidating the patterns of orthopteran evolution based on comprehensive taxon and gene sampling. Cladistics. 2015;31(6):621–51.
26. Lovejoy NR, Mullen SP, Sword GA, Chapman RF, Harrison RG. Ancient trans-Atlantic flight explains locust biogeography: molecular phylogenetics of Schistocerca. P Roy Soc B-Biol Sci. 2006;273(1588):767–74.
27. Losos JB. Convergence, adaptation, and constraint. Evolution. 2011;65(7):1827–40.
28. Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP, Clarke JA, Baker AJ, Clamp M, et al. Convergent regulatory evolution and loss of flight in paleognathous birds. Science. 2019;364(6435):74–8.
29. Vek A. Italian locust (Calliptamus italicus) in Central Kazakhstan. Proc All-Union Institute for Plant Protection. 1962;7:123–90.
30. Azhbenov VK. Outbreaks and migrations of locusts in Kazakstan. Stepnoj Bulleten. 2000;6:16–20.
31. Cao K, Mai J, Liu Q, Dou JD, Wen Q, Chen X, Xiao H, Jashenko R, Ji R. Effect of physiological factors on the flight ability of Calliptamus italicus L. Physiol Entomol. 2019; 44(3-4):230–5.
32. Sergeev MG. Ups and Downs of the Italian Locust (Calliptamus italicus L.) Populations in the Siberian Steppes: On the Horns of Dilemmas. Agronomy. 2021;11(4):746.
33. Wang X, Fang X, Yang P, Jiang X, Jiang F, Zhao D, Li B, Cui F, Wei J, Ma C, et al. The locust genome provides insight into swarm formation and long-distance flight. Nat Commun. 2014;5:2957.
34. Du B, Ding D, Ma C, Guo W. Le Kang: Locust density shapes energy metabolism and oxidative stress resulting in divergence of flight traits. Proc Natl Acad Sci. 2022;119(1): e2115753118.
35. Kimathi E, Tonnang H, Subramanian S, Cressman K, Abdel-Rahman EM, Tesfayohannes M, Niassy S, Torto B, Dubois T, Tanga CM, et al. Prediction of breeding regions for the desert locust Schistocerca gregaria in East Africa. Sci Rep-UK. 2020;10(1):11937.
36. Spottings: Aztec Spur - Throat Grasshopper Nymph. https://www.projectnoah.org. Accessed 22 Apr 2015. In.: Spotted; 2015.
37. Wikipedia. Schistocerca americana https://en.wikipedia.org/wiki/Schistocerca_americana. In.
38. CABI. Schistocerca nitens (gray bird grasshopper). https://www.cabidigitallibrary.org. Accessed 29 Aug 2012. In.; 2012.
39. Infrastructure NST. Xenocatantops brachycerus http://museum.ioz.ac.cn/species_detail.aspx?id=6474. In.
40. Zhao L, Zhang X, Qiu Z, Huang Y. De Novo Assembly and Characterization of the Xenocatantops brachycerus Transcriptome. Int J Mol Sci. 2018;19(2):520.
41. Li X, Mank JE, Ban L. The grasshopper genome reveals long-term gene content conservation of the X Chromosome and temporal variation in X Chromosome evolution. Genome Res. 2024;34(7):997–1007.
42. Gaugel SM, Hawlitschek O, Dey LS, Husemann M. Evolution of mitogenomic gene order in Orthoptera. Insect Mol Biol. 2023;32(4):387–99.
43. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90.
44. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8(8):1494–512.
45. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol Biol Evol. 2021;38(10):4647–54.
46. Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: Assessing Genomic Data Quality and Beyond. Curr Protoc. 2021;1(12): e323.

47. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658–9.
48. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150–2.
49. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics. 2011;12:124.
50. Löytynoja A. Phylogeny-aware alignment with PRANK. In: Multiple Sequence Alignment Methods. Edited by Russell DJ. Totowa, NJ: Humana Press; 2014: 155–170.
51. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006;34(Web Server issue):W609–12.
52. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37(5):1530–4.
53. Yang Z, Rannala B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol Biol Evol. 2006;23(1):212–26.
54. Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. TimeTree 5: An Expanded Resource for Species Divergence Times. Mol Biol Evol 2022;39(8):c174.
55. Álvarez-Carretero S, Kapli P, Yang Z. Beginner's Guide on the Use of PAML to Detect Positive Selection. Mol Biol Evol. 2023;40(4):41.
56. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM, et al. Gene-wide identification of episodic selection. Mol Biol Evol. 2015;32(5):1365–71.
57. Yang Z, Nielsen R. Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. Mol Biol Evol. 2000;17(1):32–43.
58. Spielman SJ, Kosakovsky PS. Relative evolutionary rate inference in HyPhy with LEISR. PeerJ. 2018;6: e4339.
59. Fukushima K, Pollock DD. Detecting macroevolutionary genotype-phenotype associations using error-corrected rates of protein convergence. Nat Ecol Evol. 2023;7(1):155–70.
60. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol Biol Evol. 2017;34(8):2115–22.
61. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021;38(12):5825–9.
62. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7.
63. Sakiyama T. Purine nucleoside phosphorylase (PNP). Nihon Rinsho. 1996;54(12):3220–5.
64. Pogosian LG, Akopian Z. Purine nucleoside phosphorylase. Biomed Khim. 2013;59(5):483–97.
65. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 1997;13(5):555–6.
66. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 2005;22(12):2472–9.
67. Kosakovsky PS, Poon A, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A, et al. HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. Mol Biol Evol. 2020;37(1):295–9.
68. Wirthlin ME, Schmid TA, Elie JE, Zhang X, Kowalczyk A, Redlich R, Shvareva VA, Rakuljic A, Ji MB, Bhat NS, et al. Vocal learning-associated convergent evolution in mammalian proteins and regulatory elements. Science. 2024;383(6690):eabn3263.
69. Sadanandan KR, Ko MC, Low GW, Gahr M, Edwards SV, Hiller M, Sackton TB, Rheindt FE, Sin S, Baldwin MW. Convergence in hearing-related genes between echolocating birds and mammals. P Natl Acad Sci USA. 2023;120(43): e1987627176.
70. Foote AD, Liu Y, Thomas GW, Vinař T, Alföldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V, et al. Convergent evolution of the genomes of marine mammals. Nat Genet. 2015;47(3):272–5.
71. Dingle H. Migration: the biology of life on the move: Oxford University Press, USA; 2014.
72. Chapman JW, Reynolds DR, Wilson K. Long-range seasonal migration in insects: mechanisms, evolutionary drivers and ecological consequences. Ecol Lett. 2015;18(3):287–302.
73. Chowdhury S, Fuller RA, Dingle H, Chapman JW, Zalucki MP. Migration in butterflies: a global overview. Biol Rev. 2021;96(4):1462–83.
74. Dingle H. Animal migration: is there a common migratory syndrome? J Ornithol. 2006;147(2):212–20.
75. Richard P, Feng S, Tsai YL, Li W, Rinchetti P, Muhith U, Irizarry-Cole J, Stolz K, Sanz LA, Hartono S, et al. SETX (senataxin), the helicase mutated in AOA2 and ALS4, functions in autophagy regulation. Autophagy. 2021;17(8):1889–906.
76. Kim HD, Choe J, Seo YS. The sen1(+) gene of Schizosaccharomyces pombe, a homologue of budding yeast SEN1, encodes an RNA and DNA helicase. Biochemistry-US. 1999;38(44):14697–710.
77. Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell. 2011;42(6):794–805.
78. Richard P, Manley JL. R Loops and Links to Human Disease. J Mol Biol. 2017;429(21):3168–80.
79. Zhang CX, Brisson JA, Xu HJ. Molecular Mechanisms of Wing Polymorphism in Insects. Annu Rev Entomol. 2019;64:297–314.
80. Bunnell TM, Burbach BJ, Shimizu Y, Ervasti JM. β-Actin specifically controls cell growth, migration, and the G-actin pool. Mol Biol Cell. 2011;22(21):4047–58.
81. Sadek KH, Cagampang FR, Bruce KD, Macklon N, Cheong Y. Variation in stability of housekeeping genes in healthy and adhesion-related mesothelium. Fertil Steril. 2012;98(4):1023–7.
82. Hou L, Guo S, Ding D, Du B, Wang X. Neuroendocrinal and molecular basis of flight performance in locusts. Cell Mol Life SCI. 2022;79(6):325.

Publisher's Note