

The Oriental Fruit Fly, *Bactrocera dorsalis*, in China: Origin and Gradual Inland Range Expansion Associated with Population Growth

Xuanwu Wan¹, Francesco Nardi², Bin Zhang¹, Yinghong Liu^{1*}

1 Chongqing Key Laboratory of Entomology and Pest Control Engineering, College of Plant Protection, Southwest University, Chongqing, China, **2** Department of Evolution Biology, University of Siena, Siena, Italy

Abstract

The oriental fruit fly, *Bactrocera dorsalis*, expanded throughout mainland China in the last century to become one of the most serious pests in the area, yet information on this process are fragmentary. Three mitochondrial genes (*nad1*, *cytb* and *nad5*) were used to infer the genetic diversity, population structure and demographic history of the oriental fruit fly from its entire distribution range in China. High levels of genetic diversity, as well as a significant correspondence between genetic and geographic distances, suggest that the invasion process might have been gradual, with no associated genetic bottlenecks. Three population groups could be identified, nevertheless the overall genetic structure was weak. The effective number of migrants between populations, estimated using the coalescent method, suggested asymmetric gene flow from the coastal region of Guangdong to most inland regions. The demographic analysis indicates the oriental fruit fly underwent a recent population expansion in the Central China. We suggest the species originated in the coastal region facing the South China Sea and gradually expanded to colonize mainland China, expanding here to high population numbers.

Citation: Wan X, Nardi F, Zhang B, Liu Y (2011) The Oriental Fruit Fly, *Bactrocera dorsalis*, in China: Origin and Gradual Inland Range Expansion Associated with Population Growth. PLoS ONE 6(10): e25238. doi:10.1371/journal.pone.0025238

Editor: Thomas Mailund, Aarhus University, Denmark

Received: June 17, 2011; **Accepted:** August 30, 2011; **Published:** October 3, 2011

Copyright: © 2011 Wan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by the National Basic Research Program of China (2009CB125903). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: yhliu@swu.edu.cn

Introduction

The oriental fruit fly, *Bactrocera dorsalis* (Hendel), is one of the most important pests on fruits and vegetables across South East Asia and the Pacific region [1]. Being highly polyphagous, the oriental fruit fly can infest a wide variety of fruit crops, such as citrus, mandarin, peach and mango [2,3], and induce significant economic losses through direct fruit damage, fruit drop and export limitations associated to quarantine restrictions. Furthermore, due to its broad host range, relatively ample climate tolerance, high reproductive potential and dispersal capacity [4], the oriental fruit fly is considered to have a high invasive potential.

The oriental fruit fly has significantly expanded its geographic distribution in the last century. Early records report its presence in 1912 in Taiwan [5]. Henceforth, this species colonized different areas of the Asian and Pacific regions, such as India, Pakistan, Nepal, Vietnam, Laos, Burma, Thailand, Sri Lanka, the Northern Mariana Islands (eradicated), Hawaii, Guam, with transient appearances in California and Florida [1,2]. The potential distribution analysis showed that the oriental fruit fly is likely to expand in the next future North and South-ward into areas currently cooler [2].

After its initial recognition in Taiwan, the oriental fruit fly was reported in 1934 on Hainan Island, China [6], and sparsely in disjointed areas of Southern China until the 1970s. Since the 1980s the population size of the oriental fruit fly increased quickly and the distribution area expanded rapidly to cover most areas south of the 26N parallel. In the last decade the oriental fruit fly

expanded across the Yangtze River to reach the 32N parallel [7–9] and is expected to expand further North [10].

Despite the economic and ecological threats associated with the invasion of the oriental fruit fly, data on its phylogeography are scarce. An early study investigated the relationship between two laboratory and three wild populations [11]. Afterwards, studies on *Bactrocera dorsalis* population genetics and phylogeography mostly addressed specific and/or geographically limited issues [12–17], until a first integrated attempt to study this species in a substantial part of its range conducted by Aketarawong et al. [18]. Aketarawong and colleagues suggested China as the origin of fly populations in South-East Asia and could describe a well supported pattern of expansion from the region of Guangdong to this latter area. Nevertheless, their study included only two Chinese populations, from Guangdong and Taiwan, limiting their possibility to study the dispersal of the oriental fruit fly in mainland China.

Mitochondrial DNA, due to its accelerated rate of evolution, short coalescence time and simple maternal inheritance, has been used as a marker of choice for historical phylogeography and complements well with the information provided by microsatellite markers. Thanks to the possibility to reconstruct evolutionary relationships among haplotypes, the mitochondrial DNA is particularly informative to reconstruct historical processes, such as the identification of the region of origin of a species, the pathways of invasion and historical demography, and has been repeatedly used to study the spread of alien species [19–26]. Furthermore, mitochondrial markers have repeatedly been used in

Table 1. Sampling locations.

Location, province	Code	n°	Longitude (E)	Latitude (N)	year
Fuzhou, Fujian	FZ	17	119°28'	26°15'	2009
Wenchang, Hainan	WC	16	110°76'	19°68'	2010
Nanning, Guangxi	NN	20	108°46'	22°78'	2010
Huaxi, Guizhou	HX	9	106°67'	26°44'	2009
Wuhan, Hubei	WH	20	114°36'	30°48'	2009
Guangzhou, Guangdong	GZ	19	113°28'	23°18'	2009
Nanchang, Jiangxi	NC	20	115°79'	28°62'	2009
Jianshui, Yunnan	JS	20	102°82'	23°70'	2010
Jiangjin, Chongqing	JJ	20	106°25'	29°08'	2009
Wanzhou, Chongqing	WZ	20	108°50'	30°75'	2009
Wulong, Chongqing	WL	20	108°97'	28°42'	2009
Xiushan, Chongqing	XS	20	107°02'	29°30'	2009

n°, number of individuals studied.

doi:10.1371/journal.pone.0025238.t001

fruit flies, such as the olive fly [27–29], the pumpkin fruit fly [30], the melon fruit fly [31] and the Mediterranean fruit fly [32–34]. Limitation to the use of mitochondrial markers is that the entire molecule is non recombining and inherited as a single locus, henceforth limiting the possibility to average across markers to take into account the random nature of the coalescence process.

With this study we specifically focus on Chinese populations, previously identified as a likely source for the species, to tackle the issue of the origin and range expansion of the oriental fruit fly. Specifically at issue are: a) a first exploration of Chinese diversity and definition of local genetic groups; b) the reconstruction of the major routes of expansion in mainland China and the definition of the demographic profile associated with the expansion, and c) the

identification of the region of origin of the species and a re-evaluation of the different hypotheses proposed in the literature.

Materials and Methods

Sampling, DNA extraction and sequencing

Samples of *Bactrocera dorsalis* were collected from 12 locations covering the entire distribution range in China during years 2009–2010 using traps baited with Methyl Eugenol (Tab. 1; Fig. 1). Specimens were preserved in 95% ethanol at -20°C until processing.

Total DNA was extracted using the DNeasy Blood and Tissue Kit (QIAGEN) from 221 individual specimens. Three fragments of the mitochondrial genome (574 bp of *nad1*, 750 bp of *cytb* and 658 bp of *nad5*) were amplified from each and all individuals using primer pairs *nad1*-F (5'-TTTAGTTGCTTGGTTGTGTAT-TCC-3')/*nad1*-R (5'-GAAAAAGGTAAAAAAGCTTTTCAAG-C-3'), *cytb*-F (5'-AACTCTTCACGCCAACGG)/*cytb*-R (5'-G-GTCGTGCTCCAATTCAT-3') and *nad5*-F (5'-TAACCCAA-TACACCTCCT-3'), *nad5*-R (5'-GGTAACTGCTGGGGTT-TA-3'). Primers for *nad1* are from Nardi et al. (2005), primers for *cytb* and *nad5* were designed based on available complete mitochondrial genome sequences of tephritids, including *B.dorsalis* [35]. Amplifications were carried out for 35 cycles of 30" at 94°C, 1' at 54°C, 1' 30" at 72°C, with an initial denaturation step of 5' at 94°C and a final extension of 10' at 72°C. Amplification products were purified and sequenced by Invitrogen Biotechnology Co. (Shanghai, China) on both strands using PCR primers. After manual correction and assembly, unique sequences were deposited in GenBank under accession numbers JF521024–JF511166 (*cytb*), JF521167–JF521298 (*nad1*) and JF521299–JF521440 (*nad5*).

Data analysis

Sequences were aligned using ClustalX (ver. 2.0) [36] and unique haplotypes were identified in ARLEQUIN (ver. 3.5) [37]. Descriptive statistics (number of variable sites, number of

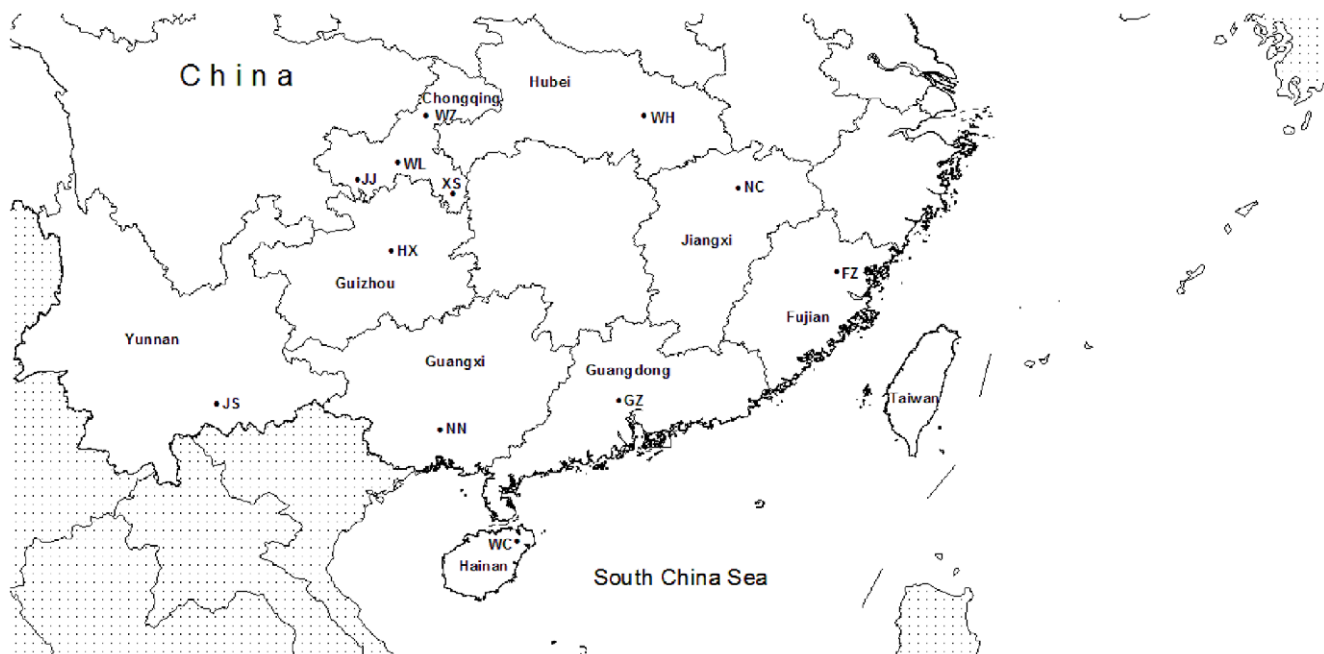


Figure 1. Collecting sites. See Tab. 1 for complete collection information.
doi:10.1371/journal.pone.0025238.g001

haplotypes, haloptype diversity, nucleotide diversity, average number of nucleotide difference between haplotypes) were calculated in Dnasp (ver. 5.0) [38].

Spatial analysis of molecular variance was performed using SAMOVA (ver. 1.0) [39] to identify population groups, with concatenated sequences of the three genes for each individual, longitude and latitude information as input data. The most supported number of groups (K) was determined by repeating the analysis with K ranging from 2 to 6 and selecting the subdivision scheme associated with highest F_{CT} . An AMOVA hierarchical analysis of variance was performed using ALEQUIN to partition total variance in its components among groups, among populations and within populations, based on the groups inferred by the SAMOVA analysis. The correlation between genetic ($F_{ST}/1-F_{ST}$) and geographic distance matrices (in ln scale) [40] was tested using the IBDWS web service [41] with 10000 randomizations. Median-joining networks of haplotypes of each of the three genes were constructed using NETWORK (ver. 4.6) [42,43] to study the evolutionary relationships among haplotypes.

The extent of gene flow between population pairs was studied using the coalescent-based strategy implemented in MIGRATE (ver. 3.2.7) [44]. To determine if there was asymmetrical gene flow between populations, the mutation scaled effective immigration rate ($M = m/\mu$) entering and leaving each population per generation and the mutation scaled effective population size ($\Theta = N_e\mu$) were jointly estimated applying the Bayesian search strategy. $N_e m$ was calculated by multiplying these latter values. Four independent runs of MIGRATE, each consisting of one long chain of 100,000,000 generations with the initial 10,000 excluded as burn-in of the analysis, were conducted to assess consistency in the results, changing seed number at each run.

The demographic history of all populations pooled together and of each of the three population groups identified by the SAMOVA analysis was studied using mismatch distributions in ARLEQUIN. Tajimas' D and Fu's F_s were calculated to test for neutrality. Population size before expansion (θ_0), population size after expansion (θ_1), population expansion time (τ), and sum of squared deviation (SSD) between observed and expected mismatch distributions were similarly calculated. All parameters were tested

against the expected values under the hypothesis of a recent population expansion based on 1000 bootstrap replicates.

Results

Genetic diversity

Collapsing of individual sequences led to the identification of 132, 143 and 142 unique haplotypes for genes *nad1*, *cytb* and *nad5*, respectively, or 164 after concatenation of the three markers for each individual.

Basic descriptive statistics, calculated for each population based on concatenated sequences as well as each of the three genes independently, are shown in Tab.S1. The number of haplotypes per population (n) ranged from 9 to 20, 6 to 19, 8 to 20, 9 to 20 in concatenated sequences, *nad1*, *cytb* and *nad5*, respectively. Haplotype diversity (H) ranged from 0.9006 to 1, 0.8889 to 1, 0.8947 to 1 and 0.8947 to 1, nucleotide diversity (π) from 0.0093 to 0.0125, 0.0077 to 0.0147, 0.0105 to 0.0194 and 0.0081 to 0.0117, similarly in concatenated sequences, *nad1*, *cytb* and *nad5*. These figures suggest that all populations retain fairly high levels of genetic diversity.

Genetic structure

Monitoring of F_{CT} values in the SAMOVA analysis suggested 3 as the optimal number of population groups ($F_{CT3} = 0.04231$). The 12 populations were clustered as follows: Jiangjin, Wulong, Wanzhou, Xiushan, Huaxi, Nanning, Wuhan, Nanchang and Jianshui; Fuzhou and Guangzhou; Wenchang. These three groups correspond to three geographically well defined regions that we refer to as Central, South-East and Southern China in the following presentation (see Fig. 1).

The AMOVA analysis revealed that a substantial portion of genetic differentiation is partitioned among groups (4.22% based on concatenated sequences, from 2.29% to 6.96% based on individual genes) and inside populations (94.95%; 92.63% to 96.89%), while genetic differentiation between populations inside each of the three groups identified is limited (0.84%; 0.4% to 1.29%) (Tab. 2). Accordingly, differentiation among groups (F_{CT}) and within populations (F_{ST}) are highly significant, while

Table 2. Partitioning of genetic variation at different hierarchical levels.

Gene analyzed	Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	Fixation indices
concatenated sequences	Among groups	2	65.720	0.48104Va	4.22	$F_{CT} = 0.04217^{**}$
	Among populations within groups	9	113.464	0.09545Vb	0.84	$F_{SC} = 0.00874$
	Within populations	209	2263.522	10.83025Vc	94.95	$F_{ST} = 0.05054^{**}$
<i>nad1</i>	Among groups	2	25.133	0.22173Va	6.96	$F_{CT} = 0.06964^{**}$
	Among populations within groups	9	28.702	0.01288Vb	0.4	$F_{SC} = 0.00435$
	Within populations	209	616.391	2.94924Vc	92.63	$F_{ST} = 0.07369^{**}$
<i>Cytb</i>	Among groups	2	20.515	0.11292Va	2.29	$F_{CT} = 0.02292^{**}$
	Among populations within groups	9	49.754	0.04054Vb	0.82	$F_{SC} = 0.00842$
	Within populations	209	997.677	4.77357Vc	96.89	$F_{ST} = 0.03115^*$
<i>nad5</i>	Among groups	2	20.110	0.14676Va	4.46	$F_{CT} = 0.04458^{**}$
	Among populations within groups	9	35.047	0.04251Vb	1.29	$F_{SC} = 0.01351^*$
	Within populations	209	648.504	3.10289Vc	94.25	$F_{ST} = 0.05749^{**}$

* $P < 0.05$,

** $P < 0.01$.

doi:10.1371/journal.pone.0025238.t002

differentiation among populations between groups (F_{SC}) was not or marginally (*nad5* only) significant (Tab. 2).

The Mantel test for correlation between genetic and geographic distances revealed a significant correlation based on the concatenated dataset ($r = 0.3667$; $P = 0.008$) (Fig.S1 A) as well as based on each individual gene ($r = 0.3852$, $P = 0.005$; $r = 0.3010$, $P = 0.025$; $r = 0.3852$, $P = 0.008$ for *nad1*, *cytb* and *nad5*, respectively) (Fig.S1 B, C, D).

MJ networks of haplotypes

Median Joining networks reconstructed from haplotypes of the three genes are shown in Fig. 2. Networks are generally star-like with limited substructure. Some haplotypes positioned in the center of the networks are found at higher frequency in two or all three population groups (such as H77 for *nad1*, H7 and H8 for *cytb*,

H4 for *nad5*), with most remaining haplotypes that are found in one single population group, generally at low frequency and connecting to central haplotypes through few mutations. Some missing haplotypes (32, 53 and 50 in *nad1*, *cytb* and *nad5*, respectively) were inferred.

Gene flow

The effective mutation scaled population size was estimated for each population and the amount of mutation scaled immigration rate in both directions was estimated for all 66 population pairs (Tab.S2). High levels of migration rate were detected among populations, ranging from 87.6 (Xiushan to Guangzhou) to 853.0 (Guangzhou to Fuzhou). Migration rate was generally symmetrical in population pairs, i.e. the immigration rate was at par with the emigration rate. Some instances of asymmetric migration rate

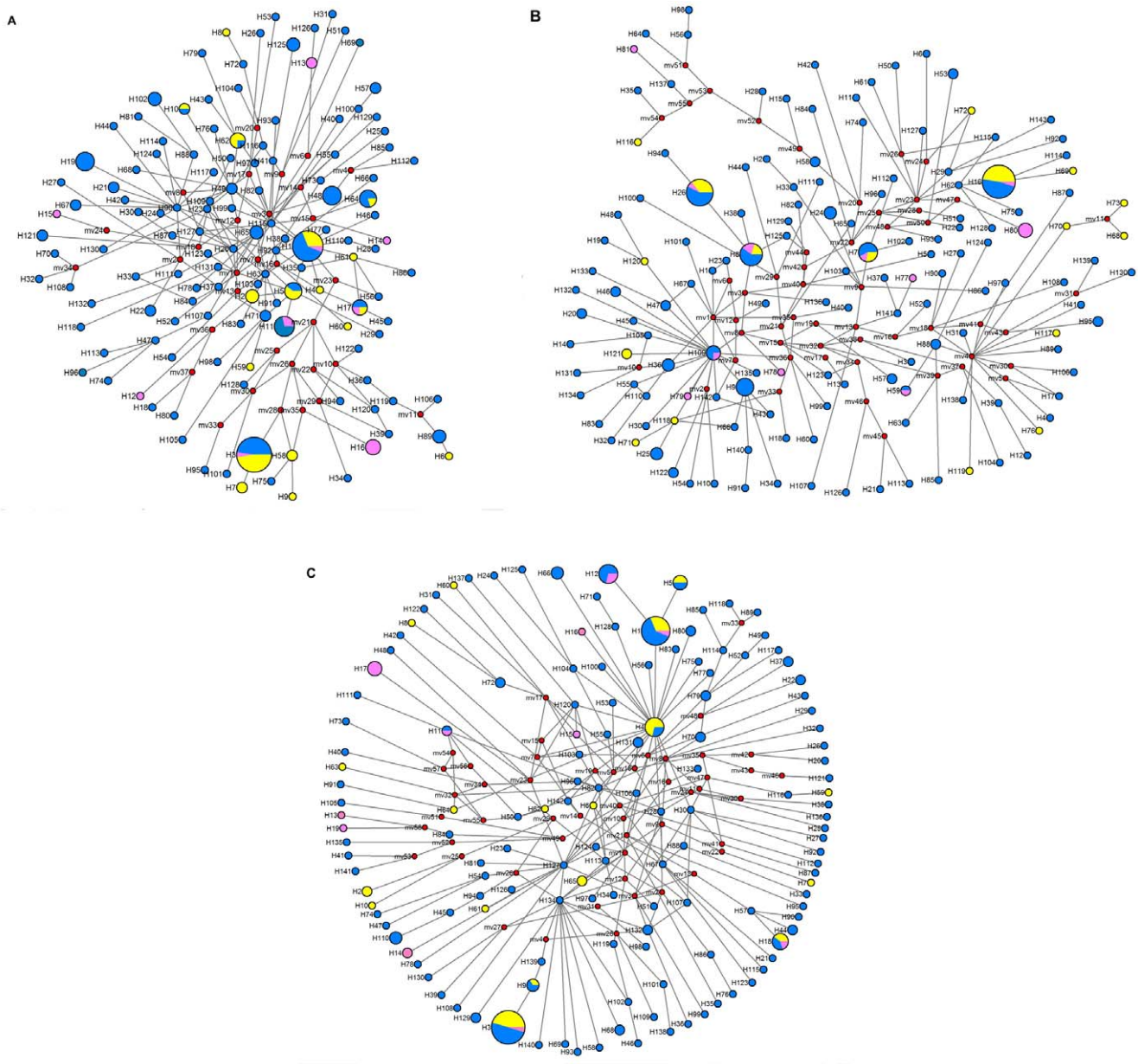


Figure 2. Median joining networks of haplotyps. Pie area proportional to haplotype frequency. A: *nad1*; B: *cytb*; C: *nad5*; blue: Central China group; yellow: South-East China group; pink: Southern China group; red: inferred haplotypes. doi:10.1371/journal.pone.0025238.g002

Table 3. Parameters of demographic history of the collated 12 populations and each of three population groups independently.

Gene	Group	θ_0	θ_1	τ	D	F_s	SSD
Conc. sequences	All	0.007	89.575	24.453	-2.1644**	-23.6507**	0.0016
	South-East	0.004	83.203	25.363	-0.8490	-0.8565	0.0177*
	South	0.064	96.464	26.805	2.1213	-0.7214	0.0353*
	Central	0.019	99.248	22.164	-2.2152**	-23.7562**	0.00170
nad1	All	1.368	46.094	5.340	-1.9658**	-24.8111**	0.0013
	South-East	0.002	10.996	7.660	-0.7696	-4.1671	0.1387
	South	0.000	41.133	10.320	-0.5318	-1.0680	0.0586*
	Central	0.991	85.156	5.236	-2.0381**	-24.9918**	0.0009
Cytb	All	0.009	63.047	10.311	-2.1575**	-24.188**	0.0011
	South-East	0.007	40.000	11.129	-0.8292	-2.9989	0.1806
	South	0.000	66.719	9.666	-0.0260	-0.6813	0.0449*
	Central	0.012	59.021	10.271	-2.1945**	-24.2877**	0.0008
nad5	All	0.011	78.594	7.656	-2.1832**	-24.7418**	0.0006
	South-East	0.000	29.727	7.928	-0.7925	-3.7418	0.0387**
	South	0.002	59.336	7.424	-0.88428	-1.6088	0.0196
	Central	0.000	99999	5.744	-2.2348**	-24.9026**	0.0294

θ_0 : effective population size before expansion; θ_1 : effective population size after expansion; τ : population expansion time; D : Tajima's D ; F_s : Fu's F_s ; SSD: sum of squared deviations between observed and expected mismatch distribution under a sudden expansion model;

* $P < 0.05$;

** $P < 0.01$.

doi:10.1371/journal.pone.0025238.t003

were nevertheless observed, based on non overlapping 95% HPD intervals for M in the two directions, namely from Guangzhou to Fuzhou, Wenchang, Wuhan, Jiangjin, Wanzhou and Xiushan. Expressed in terms of N_{em} , average gene flow inside population groups is 29.442 (range 0.778–51.673) and across population groups is 17.370 (0.322–56.664), i.e. high levels of overall gene flow are observed.

Demographic history

Significantly negative values of Tajimas' D (-2.1644 , $P = 0.0004$) and Fu's F_s (-23.6507 , $P = 0.0056$) based on concatenated sequences as well as on individual genes (see Tab. 3) indicated that the whole set of *B. dorsalis* samples studied here did not fit a simple model of neutral evolution. These same estimators calculated for the three population groups indicate that South and South-East China have values of D and F_s compatible with neutrality, while Central China has values of D and F_s significantly deviating from neutrality (Tab. 3), suggesting Central China as the responsible for the overall disequilibrium.

The mismatch distribution of all 12 *B. dorsalis* populations pooled together as well as of the Central China group only were distinctly unimodal (Fig. 3, 4), suggesting the further testing of a sudden expansion model.

The mismatch distribution of the 12 populations pooled together was compatible with the sudden expansion model based on concatenated sequences ($P_{SSD} = 0.556$) as well as each individual gene (see Tab. 3). Parameters of the expansion model were $\theta_0 = 0.007$, $\theta_1 = 89.575$ and $\tau = 24.453$. The mismatch distribution of the Central China group was similarly compatible with the sudden expansion model based on concatenated sequences ($P_{SSD} = 0.455$) and individual genes (see Tab. 3). Parameters of the expansion model for the Central China group were $\theta_0 = 0.019$, $\theta_1 = 99.248$ and $\tau = 22.164$. The sudden expansion model was in turn rejected for population groups South and South-East China ($P < 0.05$). Individual genes (see

Tab. 3) are in accord with concatenated sequences, all being compatible with the sudden expansion model for all 12 populations pooled and the Central China group ($P > 0.05$) and most rejecting the expansion model in the other two population groups.

The ratio between estimated effective population size after expansion (θ_1) and before expansion (θ_0), an estimate of the extent of population growth, is $12796 \times$ for the entire dataset and $5223 \times$ for the Central China group based on concatenated sequences. Individual genes, although with differing numerical estimates, similarly suggest a large population growth, from $33 \times$ to $7144 \times$ for the entire dataset and from $85 \times$ to actually infinite for the Central China group.

Discussion

High level of genetic diversity

Invasive species are generally associated to a loss in genetic diversity that can take place, during an invasion process, due to a) increased genetic drift associated to the temporarily reduced population size in founding colonies [45–47] and b) increased selection pressure encountered during the colonization of new habitats [48].

Chinese populations of the oriental fruit fly, nevertheless, seem to retain fairly high levels of genetic diversity, as exemplified by the high observed values of haplotype diversity. Noteworthy, this is observed even in populations such as Wuhan, Xiushan, Wulong, Wanzhou and Jiangjin that, based on collection records, established not longer than a decade ago and that display values of genetic variability actually higher than some of the oldest populations in the study set, Fuzhou and Guangzhou.

Nevertheless, some characteristics of the oriental fruit fly and the ecology of the area have to be considered that could help explain why no loss in genetic variability is observed concurrent to the colonization process. Due to the high reproductive potential of

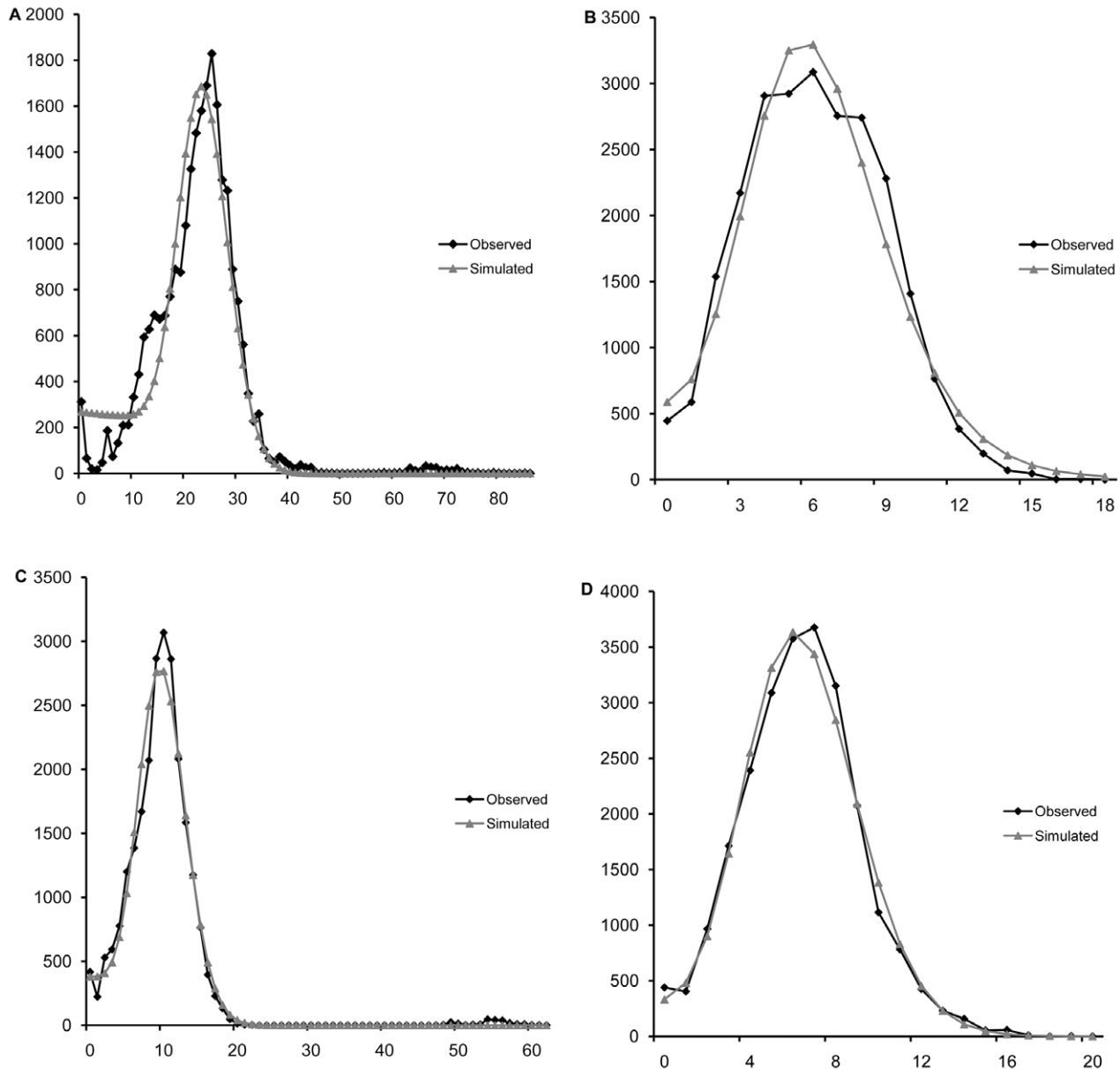


Figure 3. Observed and simulated mismatch distributions of entire sample. A, B, C and D are for concatenated sequences, *nad1*, *cytb* and *nad5*, respectively, the horizontal axis represents the number of pairwise differences, the vertical axis represents the relative frequency. doi:10.1371/journal.pone.0025238.g003

this species and high dispersal capabilities, together with the relative absence of major geographical or ecological barriers to dispersal in the area, it is possible to envision how the oriental fruit fly might have gradually expanded without any significant or prolonged bottleneck. Furthermore, the relative abundance of suitable host fruits such as mango, carambola and guava, the large orange plantations in South China and Yangtze valley, and the relative uniformity and stability of a suitable tropical and subtropical monsoon climate in this area might have exerted little novel selective pressures during the range expansion in Central China.

Furthermore, multiple introductions, or colonization of a given area from multiple sources, can counteract the drop in genetic variability associated with colonization or rescue a species from an actual loss in genetic diversity [48]. Such cases of increased genetic variability due to a secondary mixing between previously diversified populations have been repeatedly described in the

literature [49–53] and multiple introductions have been previously suggested to explain the relatively high genetic variability of the oriental fruit fly in the Yunnan province of China [13].

The oriental fruit fly may have, therefore, expanded quickly but gradually in mainland China, without the significant bottlenecks generally associated with a range expansion that takes place through invasive propagules of few individuals. This view is further in line with the high levels of gene flow and limited population differentiation observed (see below) and with the situation described by Aketarawong et al. [18] in South-East Asia and Shi [13] in the province of Yunnan.

Weak genetic structure

Although some structure could be identified by the SAMOVA analysis, leading to the partitioning of the 12 populations studied in the three groups Central, South-East and Southern China, the

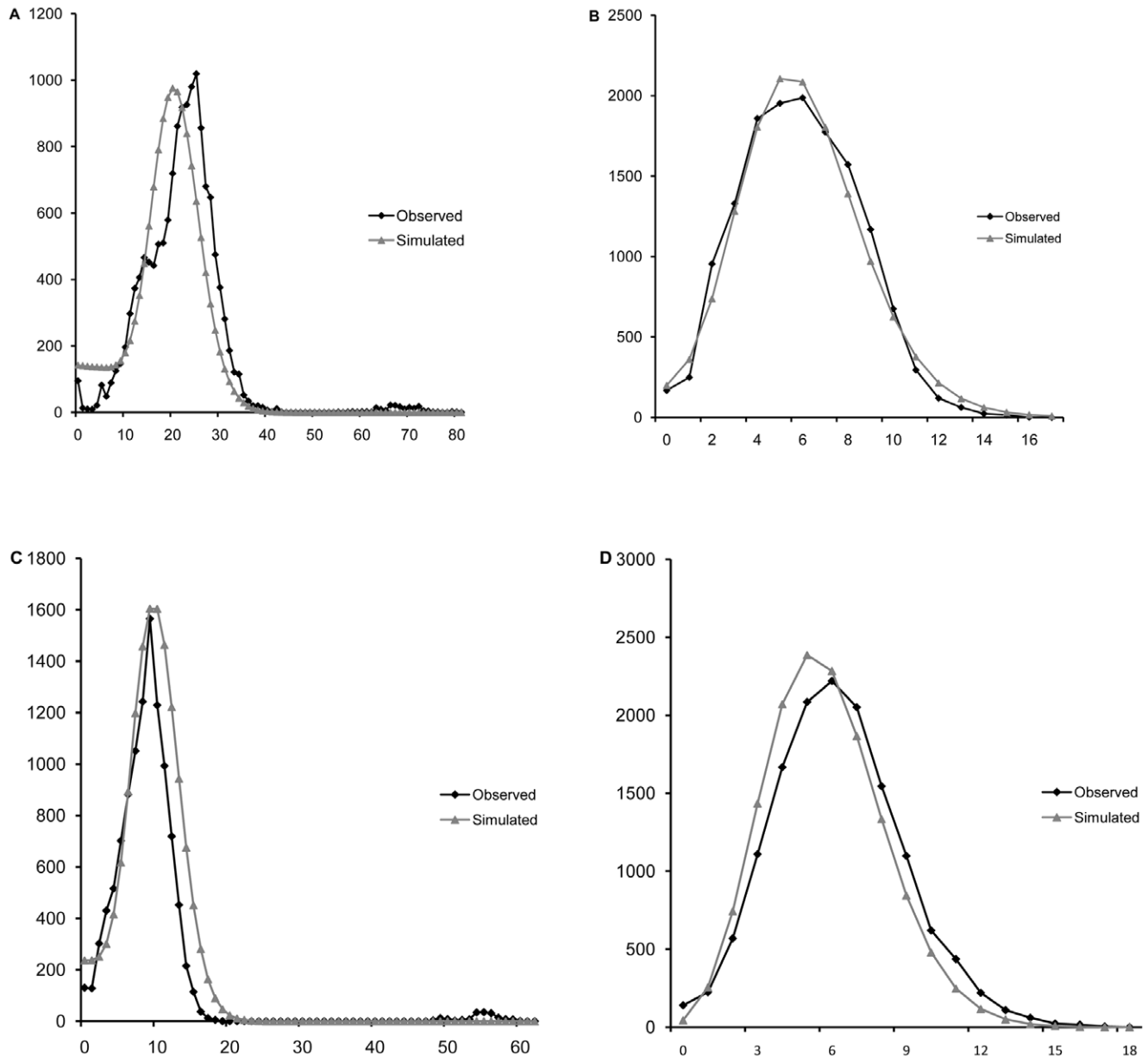


Figure 4. Observed and simulated mismatch distributions of Central China group only. A, B, C and D are for concatenated sequences, *nad1*, *cytb* and *nad5*, respectively, the horizontal axis represents the number of pairwise differences, the vertical axis represents the relative frequency. doi:10.1371/journal.pone.0025238.g004

overall level of differentiation was low. According to the analysis of molecular variance, more than 90% of variability was observed inside populations, with no or marginal differentiation between populations and some differences arising only between population groups. This is rather unexpected given the distances involved, with the area occupied by the Central group being roughly equivalent to Central Europe in size, but in line with other examples of highly vagile fruit fly species that have expanded in large areas in relatively recent times [28,54].

The median joining networks for the three genes, accordingly, did not describe any underlying structure, as networks are distinctively star-like and haplotypes sampled in the three regions, not to mention populations, appear randomly distributed. The Mantel test identified a significant, though not strong, correlation between genetic and geographic distance, suggesting a certain degree of isolation by distance with no major discontinuity.

Taken together, these observations suggest a relative genetic uniformity of the oriental fruit fly in China. This is in line with the notion that the species has high dispersal capacities, with eggs and larvae that can disperse efficiently inside host fruits both under natural conditions (i.e. along rivers and ocean currents [55]) and artificially through trades of infested fruits, and adults that can fly as far as 46 km under experimental conditions [56].

It is therefore possible to hypothesize that the natural barriers present in the study area (Daba Mountains, Hengduan Mountains, Wuyi Mountains, Yangtze River, Zhujiang River and Qiongzhou Strait) are not sufficient to interrupt or determine a significant limitation to gene flow. A similar situation has been described in the region of Yunnan, where weak genetic structure was observed among 14 oriental fruit fly populations despite the presence of three big rivers and mountain ranges running North-West to South-East in the area [57]. Opposite results, i.e. the insurgence of

a clear genetic structure, were in turn obtained for more sedentary species, such as the rice stem borer *Chilo suppressalis* [58], where rivers and mountain ranges act as effective barriers to gene flow.

Demographic history

The significant negative Tajima's D and Fu's F_s values showed that the oriental fruit fly populations in mainland China do not fit a simple model of selective neutrality [59,60] due to an excess in low frequency alleles. This, together with collection data that indicate a substantial increase of oriental fruit fly in terms of both geographic range and population numbers, suggested the possibility of a recent population expansion. In turn, analysis of Tajima's D and Fu's F_s estimators separately in the three groups indicated that the Central China group may have specifically undergone a regime of population expansion. The unimodal mismatch distribution [61] for all genes in the entire dataset and specifically in the Central China group, as well as non significant SSD values against the null hypothesis of a sudden population expansion further support this hypothesis [62]. The notion that the species underwent a significant population expansion is further in line with the observation that Median Joining networks for the three genes have a distinctly star like structure, typical of expansion demographic processes.

These results are concordant with trapping data in the area. Oldest presence of the species is reported from the area facing the island of Taiwan, where the species was first detected, and the island of Hainan. These two areas correspond to population groups South-East and South China, respectively, that based on our data show no sign of population expansion. In turn, the marked range expansion observed in the last few decades mainly interested mainland China up to the 32N parallel, a very large region corresponding to population group Central China, for which distinctive signs of population expansion could be described.

Based on historical records and the genetic data presented here, it is therefore possible to reconstruct a process of fast and recent range expansion from the coastline area facing the South China sea, where the species has been likely established for a longer period of time, to a very large mainland region that has been colonized in the last few decades. Based on the high genetic diversity in the area and limited genetic structure (see above), we hypothesize that this process of colonization may be interpreted as a gradual, though fast, range expansion associated with high population numbers and population growth. The opposite scenario of a stepping stone model of repeated colonization events through numerically limited propagules may on the other hand be excluded, as no trace of genetic bottlenecks and related drop in genetic variability is observed.

The notion that local natural barriers seem to be rather ineffective in counteracting oriental fruit fly dispersal, as well as the potential distribution analysis, suggests that the process described here of population expansion in mainland China may interest other regions north of the 32N parallel in the next decades.

Region of origin

Different hypotheses have been proposed about the geographic origin of the oriental fruit fly. While the current distributional range of the species is rather ample, from India to Hawaii encompassing all South-East Asia, historical records clearly show that marginal populations represent recent introductions, with the species being most likely East-Asian in origin. One initial

hypothesis on the geographic origin of the oriental fruit fly, based on the earliest records of its presence in the island, is Taiwan [6]. Subsequent studies, based on the high levels of local genetic variability, hypothesized Yunnan as a possible source area for the species [13], or at least an area of old colonization [57]. The most comprehensive study to date on the phylogeography of this species [18], in turn, suggested China as the source area for South-East Asian populations, and possibly for the species as a whole, based on the observation that China is equally divergent from Pacific and South-East Asian populations. Furthermore, a well supported pattern of a West-ward migration from China to colonize South-East Asia was described. Noteworthy, Aketarawong and colleagues, based on asymmetrical gene flow values, could exclude Taiwan as the source of the invasion.

Based on our extended sampling in mainland China, that greatly expands the current knowledge on local oriental fruit fly diversity beyond the region of Yunnan [13,57] and one single population in Guangdong [18], we could further explore the patterns of diversity and range expansion in China to support an origin of the species in the coastal region facing the South China Sea, corresponding in genetic terms to the population group here described as South-East China. Main support for this hypothesis is found in the asymmetric gene flow observed from location Guangzhou (Guangdong) to 6 out of 11 alternative locations scattered in mainland China. Furthermore, this hypothesis is in line with the historical records showing that population groups South and South-East China occupy regions where the species has been present since the last century, while population group Central China, that we associate with the expansion process, occupies the mainland regions where the oriental fruit fly has been reported in the last two decades only.

Further support for the coastal region of Southern China being the source area for the species as a whole comes from a joint interpretation of our results and those presented by Aketarawong et al. [18]. The two studies, taken together, cover the majority of the oriental fruit fly geographic range, and all regions that are plausible candidates for the early differentiation of the species: China and South-East Asia from Bangladesh to Hawaii. In turn, both studies describe a similar pattern of expansion from a single and the same region (Guangdong) throughout South-East Asia [18] and mainland China (this study), with a distinct signal of asymmetrical gene flow out of this region.

Supporting Information

Figure S1 Scatter plots of genetic distance vs. geographic distance (in ln scale) for pairwise population comparisons. A: concatenated sequences, B: *nad1*, C: *cytb*, D: *nad5*. (TIF)

Table S1 Genetic diversity indices. V , number of variable sites; n , number of unique haplotypes; H , haplotype diversity; π , nucleotide diversity; k , average number of nucleotide differences. (DOC)

Table S2 Estimates of population size and effective immigration rate between populations pairs. Θ : mutation scaled effective population size; M : mutation scaled effective immigration rate. In parentheses the 95% HPD intervals. Instances of asymmetrical gene flow are indicated in bold. The source population is indicated in columns, the target population in row. (DOC)

Acknowledgments

We are grateful to Xiaozhen Li, Daochao Jin, Changying Niu, Nannan Zhang and Yanbiao He for sample collections.

References

- Clarke AR, Armstrong KF, Carmichael AE, Milne JR, Raghu S, et al. (2005) Invasive phytophagous pests arising through a recent tropical evolutionary radiation: the *Bactrocera dorsalis* complex of fruit flies. *Annual Review of Entomology* 50: 293–319.
- Stephens AEA, Kriticos DJ, Leriche A (2007) The current and future potential geographical distribution of the oriental fruit fly, *Bactrocera dorsalis* (Diptera: Tephritidae). *Bulletin of Entomological Research* 97: 369–378.
- Jin T, Zeng L, Lin YY, Lu YY, Liang GW (2011) Insecticide resistance of the oriental fruit fly, *Bactrocera dorsalis* (Hendel) (Diptera: Tephritidae), in mainland China. *Pest Management Science* 67: 370–376.
- Peterson MA, Denno RF (1998) The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. *The American Naturalist* 152: 428–446.
- Hardy DE (1973) *Pacific Insects Monograph* 31: The fruit flies (Tephritidae-Diptera) of Thailand and Bordering Countries. Honolulu: Entomology Department, Bernice P. Bishop Museum. pp 1–353.
- Wang XJ (1996) The fruit flies (Diptera: Tephritidae) of the East Asian Region. *Acta Zootaxonomica Sinica* 21 (Supplement): 53.
- Yuan M, Wang B, Song CB, Rong XL, Yin Y (2008) Effect of Climate Factors and Host Plants on Population Dynamics of *Bactrocera dorsalis* (Hendel) in Suzhou. *Journal of Anhui Agriculture Science* 22: 9619–9621.
- Wang SQ, Xiang ZJ, Liu YM, Peng LC, Du G, et al. (2009) The risk analysis of the oriental fruit fly, *Bactrocera dorsalis*, in Hubei. *Plant Protection of Hubei* 3: 5–6.
- Chen LG, Zhu CG, Xia XN (2010) A preliminary study on the biological characters of *Bactrocera dorsalis* (Hendel) in Shanghai area. *Acta Agriculturae Shanghai* 1: 99–101.
- Zhou GL, Chen C, Ye J, Hu BS, Liu FQ (2007) Predicting potential ecological distribution of *Bactrocera dorsalis* in China using GARP ecological niche modeling. *Acta Ecologica Sinica* 8: 3362–3369.
- He M, Haymer DS (1997) Polymorphic intron sequences detected within and between populations of the oriental fruit fly (Diptera: Tephritidae). *Annals of Entomological Society of America* 6: 825–831.
- Dai SM, Lin CC, Chang C (2004) Polymorphic microsatellite DNA markers from the oriental fruit fly *Bactrocera dorsalis* (Hendel). *Molecular Ecology Notes* 4: 629–631.
- Shi W, Kerdelhue C, Ye H (2005) Population genetics of the oriental fruit fly, *Bactrocera dorsalis* (Diptera: Tephritidae), in Yunnan (China) based on mitochondrial DNA sequences. *Environmental Entomology* 34: 977–983.
- Aketarawong N, Bonizzoni M, Malacrida AR, Gasperi G, Thanaphum S (2006) Seventeen novel microsatellite markers from an enriched library of the pest species *Bactrocera dorsalis* sensu stricto. *Molecular Ecology Notes* 6: 1138–1140.
- Liu JH, Shi W, Ye H (2007) Population genetics analysis of the origin of the oriental fruit fly, *Bactrocera dorsalis* Hendel (Diptera: Tephritidae), in northern Yunnan Province, China. *Entomological Science* 10: 11–19.
- Muraji M, Nakahara S, Ishida T, Minoura K, Miyazaki I, et al. (2008) The Philippines is a possible source of the *Bactrocera dorsalis* complex species (Diptera, Tephritidae) occasionally collected in the Ryukyu Islands of Japan: analyses of mitochondrial DNA. *Applied Entomology and Zoology* 4: 609–615.
- Nakahara S, Kobashigawa Y, Muraji M (2008) Genetic variations among and within populations of the oriental fruit fly, *Bactrocera dorsalis* (Diptera; Tephritidae), detected by PCR-RFLP of the mitochondrial control region. *Applied Entomology and Zoology* 3: 457–465.
- Aketarawong N, Bonizzoni M, Thanaphum S, Gomulski LM, Gasperi G, et al. (2007) Inferences on the population structure and colonization process of the invasive oriental fruit fly, *Bactrocera dorsalis* (Hendel). *Molecular Ecology* 16: 3522–3532.
- Finn DS, Theobald DM, Black IV, Poff NL (2006) Spatial population genetic structure and limited dispersal in a Rocky Mountain alpine stream insect. *Molecular Ecology* 15: 3553–3566.
- Venkatesan M, Westbrook CJ, Hauer MC, Rasgon JL (2007) Evidence for a population expansion in the West Nile virus vector *Culex tarsalis*. *Molecular Biology and Evolution* 5: 1208–1218.
- Ciosi M, Miller NJ, Kim KS, Giordano R, Estoup A, et al. (2008) Invasion of Europe by the western corn rootworm, *Diabrotica virgifera virgifera*: multiple transatlantic introductions with various reductions of genetic diversity. *Molecular Ecology* 17: 3614–3627.
- Miller NJ, Dorhout DL, Rice ME, Sappington TW (2009) Mitochondrial DNA variation and range expansion in western bean cutworm (Lepidoptera: Noctuidae): no evidence for a recent population bottleneck. *Environmental Entomology* 1: 274–280.
- Suárez NM, Betancor E, Klassert TE, Almeida T, Hernández M (2009) Phylogeography and genetic structure of the Canarian common chaffinch (*Fringilla coelebs*) inferred with mtDNA and microsatellite loci. *Molecular Phylogenetics Evolution* 2: 556–564.

Author Contributions

Conceived and designed the experiments: XW BZ YL. Performed the experiments: XW. Analyzed the data: XW FN. Wrote the paper: XW FN YL BZ.

- Bryja J, Smith C, Konečný A, Reichaid M (2010) Range-wide population genetic structure of the European bitterling (*Rhodeus amarus*) based on microsatellite and mitochondrial DNA analysis. *Molecular Ecology* 19: 4708–4722.
- Siedchlag AC, Benozzati ML, Passoni JC, Rodrigues MT (2010) Genetic structure, phylogeny, and biogeography of Brazilian eyelid-less lizards of genera *Calyptommatus* and *Nothobachia* (*Squamata, Gymnophthalmidae*) as inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* 2: 622–630.
- Zepeda-Paulo FA, Simon JC, Ramirez CC, Fuentes-Contreras E, Margaritopoulos JT, et al. (2010) The invasion route for an insect pest species: the tobacco aphid in the New World. *Molecular Ecology* 21: 4738–4752.
- Nardi F, Carapelli A, Dallai R, Frati F (2003) The mitochondrial genome of the olive fly *Bactrocera oleae*: two haplotypes from distant geographical locations. *Insect Molecular Biology* 12: 605–611.
- Nardi F, Carapelli A, Dallai R, Roderick GK, Frati F (2005) Population structure and colonization history of the olive fly *Bactrocera oleae* (Diptera, Tephritidae). *Molecular Ecology* 14: 2729–2738.
- Nardi F, Carapelli A, Boore JL, Roderick GK, Dallai R, et al. (2010) Domestication of olive fly through a multi-regional host shift to cultivated olives: Comparative dating using complete mitochondrial genomes. *Molecular Phylogenetics and Evolution* 2: 678–686.
- Mun J, Bohonak AJ, Roderick GK (2003) Population structure of the pumpkin fruit fly *Bactrocera depressa* (Tephritidae) in Korea and Japan: Pliocene allopatry or recent invasion? *Molecular Ecology* 12: 2941–2951.
- Hu J, Zhang IJ, Nardi F, Zhang RJ (2008) Population genetic structure of the melon fly, *Bactrocera cucurbitae* (Diptera: Tephritidae), from China and Southeast Asia. *Genetica* 134: 319–324.
- Meixner MD, Mepheron BA, Silva JG, Gasparich GE, Sheppard WS (2002) The Mediterranean fruit fly in California: evidence for multiple introductions and persistent populations based on microsatellite and mitochondrial DNA variability. *Molecular Ecology* 11: 891–899.
- Alaoui A, Imoulan A, EL Alaoui-Talibi Z, EL Meziane A (2010) Genetic structure of Mediterranean fruit fly (*Ceratitidis capitata*) populations from Moroccan endemic forest of *Argania spinosa*. *International Journal of Agriculture & Biology* 2: 291–298.
- Elfékih S, Makni M, Haymer DS (2010) Detection of novel mitochondrial haplotype variants in populations of the Mediterranean fruit fly, *Ceratitidis capitata*, from Tunisia, Israel and Morocco. *Journal of Applied Entomology* 8: 647–651.
- Yu DJ, Xu L, Nardi F, Li JG, Zhang RJ (2007) The complete nucleotide sequence of the mitochondrial genome of the oriental fruit fly, *Bactrocera dorsalis* (Diptera: Tephritidae). *Gene* 396: 66–74.
- Larkin MA, Blackshields G, Chenna R, McGettigan PA, McWilliam H, et al. (2007) Clustal W and clustal X version 2.0. *Bioinformatics* 21: 2947–2948.
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resour* 3: 564–567.
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
- Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* 12: 2571–2581.
- Mantel N (1967) The detection of disease clustering as a generalized regression approach. *Cancer Research* 27: 209–220.
- Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *BMC Genetics* 6: 13.
- Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37–48.
- Polzin T, Daneschmand SV (2003) On Steiner trees and minimum spanning trees in hypergraphs. *Operations Research Letters* 31: 12–20.
- Berli P (2006) Comparison of Bayesian and maximum likelihood inference of population genetic parameters. *Bioinformatics* 3: 341–345.
- Baker DA, Loxdale HD, Edwards OR (2003) Genetic variation and founder effects in the parasitoid wasp, *Diaeretiella rapae* (M'Intosh) (Hymenoptera: Braconidae: Aphidiidae), affecting its potential as a biological control agent. *Molecular Ecology* 12: 3303–3311.
- Schmitt T, Cizek D, Kovicka M (2005) Genetics of a butterfly relocation: large, small and introduced populations of the mountain endemic *Erebia ephiphron silesiana*. *Biological Conservation* 123: 11–18.
- Grapputo A, Bisazza A, Pilastro A (2006) Invasive success despite reduction of genetic diversity in the European populations of eastern mosquitofish (*Gambusia holbrooki*). *Italian Journal of Zoology* 73: 67–73.
- Suarez AN, Tsutsui ND (2008) The evolutionary consequences of biological invasions. *Molecular Ecology* 17: 351–360.
- Daehler CC, Strong DR (1997) Hybridization between introduced smooth cordgrass (*Spartina alterniflora*; Poaceae) and native California cordgrass (*S. foliosa*)

- in San Francisco Bay, California, USA. *American Journal of Botany* 84: 607–611.
50. Anttila CK, King RA, Ferris C, Ayres DR, Strong DR (2000) Reciprocal hybrid formation of *Spartina*, San Francisco Bay. *Molecular Ecology* 9: 765–770.
 51. Ayres DR, Smith DL, Zaremba K, Klohr S, Strong DR (2004) Spread of exotic cordgrasses and hybrids (*Spartina sp.*) in the tidal marshes of San Francisco Bay, California, USA. *Biology Invasions* 6: 221–231.
 52. Ellstrand NC, Schierenbeck KA (2006) Hybridization as a stimulus for the evolution of invasiveness in plants? *Euphytica* 1: 35–46.
 53. Whitfield CW, Behura SK, Berlocher SH, Clark AG, Johnston JS, et al. (2006) Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera*. *Science* 314: 642–645.
 54. Augustinos AA, Mamuris Z, Stratikopoulos EE, D'Amelio S, Zacharopoulou A, et al. (2005) Microsatellite analysis of olive fly populations in the Mediterranean indicates a westward expansion of the species. *Genetica* 125: 231–241.
 55. Chen P, Ye H, Mu QA (2007) Migration and dispersal of the oriental fruit fly, *Bactrocera dorsalis* in regions of Nujiang River based on fluorescence mark. *Acta Ecologica Sinica* 6: 2468–2476.
 56. Liang F, Wu JJ, Liang GQ (2001) The first report of the test on the flight ability of oriental fruit fly. *Acta Agriculture University Jiangxi* 2: 259–260.
 57. Shi W, Kerdelhué C, Hui Y (2010) Population genetic structure of the oriental fruit fly, *Bactrocera dorsalis* (Hendel) (Diptera: Tephritidae) from Yunnan province (China) and nearby sites across the border. *Genetica* 138: 377–385.
 58. Meng XF, Shi M, Chen XX (2008) Population genetic structure of *Chilo suppressalis* (Walker) (Lepidoptera: Crambidae): strong subdivision in China inferred from microsatellite markers and mtDNA gene sequences. *Molecular Ecology* 17: 2880–2897.
 59. Ohta T, Gilespie JH (1996) Development of neutral and nearly neutral theories. *Theoretical Population Biology* 49: 128–142.
 60. Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, et al. (1998) Genetic traces of ancient demography. *Proceedings of the National Academic Sciences of the United States of America* 95: 1691–1697.
 61. Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* 9: 552–569.
 62. Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. *Evolution* 49: 608–615.