

RESEARCH ARTICLE

Open Access

# Sequence analysis of mitochondrial *ND1* gene can reveal the genetic structure and origin of *Bactrocera dorsalis* s.s.

Zhong-Zhen Wu<sup>1</sup>, Hong-Mei Li<sup>2</sup>, Shu-Ying Bin<sup>1</sup>, Jun Ma<sup>3</sup>, Hua-Liang He<sup>1</sup>, Xian-Feng Li<sup>1</sup>, Fei-Liang Gong<sup>1</sup> and Jin-Tian Lin<sup>1\*</sup>

## Abstract

**Background:** The oriental fruit fly, *Bactrocera dorsalis* s.s., is one of the most important quarantine pests in many countries, including China. Although the oriental fruit fly has been investigated extensively, its origins and genetic structure remain disputed. In this study, the NADH dehydrogenase subunit 1 (*ND1*) gene was used as a genetic marker to examine the genetic diversity, population structure, and gene flow of *B. dorsalis* s.s. throughout its range in China and southeast Asia.

**Results:** Haplotype networks and phylogenetic analysis indicated two distinguishable lineages of the fly population but provided no strong support for geographical subdivision in *B. philippinensis*. Demographic analysis revealed rapid expansion of *B. dorsalis* s.s. populations in China and Southeast Asia in the recent years. The greatest amount of genetic diversity was observed in Manila, Pattaya, and Bangkok, and asymmetric migration patterns were observed in different parts of China. The data collected here further show that *B. dorsalis* s.s. in Yunnan, Guangdong, and Fujian Provinces, and in Taiwan might have different origins within southeast Asia.

**Conclusions:** Using the mitochondrial *ND1* gene, the results of the present study showed *B. dorsalis* s.s. from different parts of China to have different genetic structures and origins. *B. dorsalis* s.s. in China and southeast Asia was found to have experienced rapid expansion in recent years. Data further support the existence of two distinguishable lineages of *B. dorsalis* s.s. in China and indicate genetic diversity and gene flow from multiple origins.

The sequences in this paper have been deposited in GenBank/NCBI under accession numbers KC413034–KC413367.

**Keywords:** Genetic structure, Origin, *Bactrocera dorsalis* s.s., Mitochondrial *ND1*

## Background

The *Bactrocera dorsalis* (Diptera: Tephritidae) is a major horticultural pest. More than 70 species have been identified within this group [1]. Some of the *B. dorsalis* species are highly destructive because of their wide host-ranges, considerable ecological adaptability, pronounced reproductive potential, and dispersal capacity. The *B. dorsalis* s.s., *B. papayae*, and *B. philippinensis* species have become invasive in many parts of China. Despite previous studies that have described the subtle morphological characteristics distinguishing these three species [1-3], taxonomists

still doubt whether *B. dorsalis* s.s., *B. papayae*, and *B. philippinensis* are truly separate species [4,5].

*B. dorsalis* s.s. was first recorded in Taiwan in 1912 and is now widely distributed in most countries in the Asia-Pacific region [6]. Specifically, it has dispersed from China to the northern parts of the Indian subcontinent over the past 90 years [1]. *B. dorsalis* s.s. has been observed multiple times in China during the past hundred years [7-9]. Currently, *B. dorsalis* s.s. is distributed mostly in China's southern provinces [10].

Mitochondrial DNA analyses have been conducted to study the genetic structure and origin of *B. dorsalis* s.s. populations in China and Southeast Asia [11-15]. However, the results of these studies have been inconsistent, although it is not known whether these inconsistencies

\* Correspondence: studywell888@gmail.com

<sup>1</sup>Institute for Management of Invasive Alien Species, 314 Yingdong teaching building, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, PR China

Full list of author information is available at the end of the article

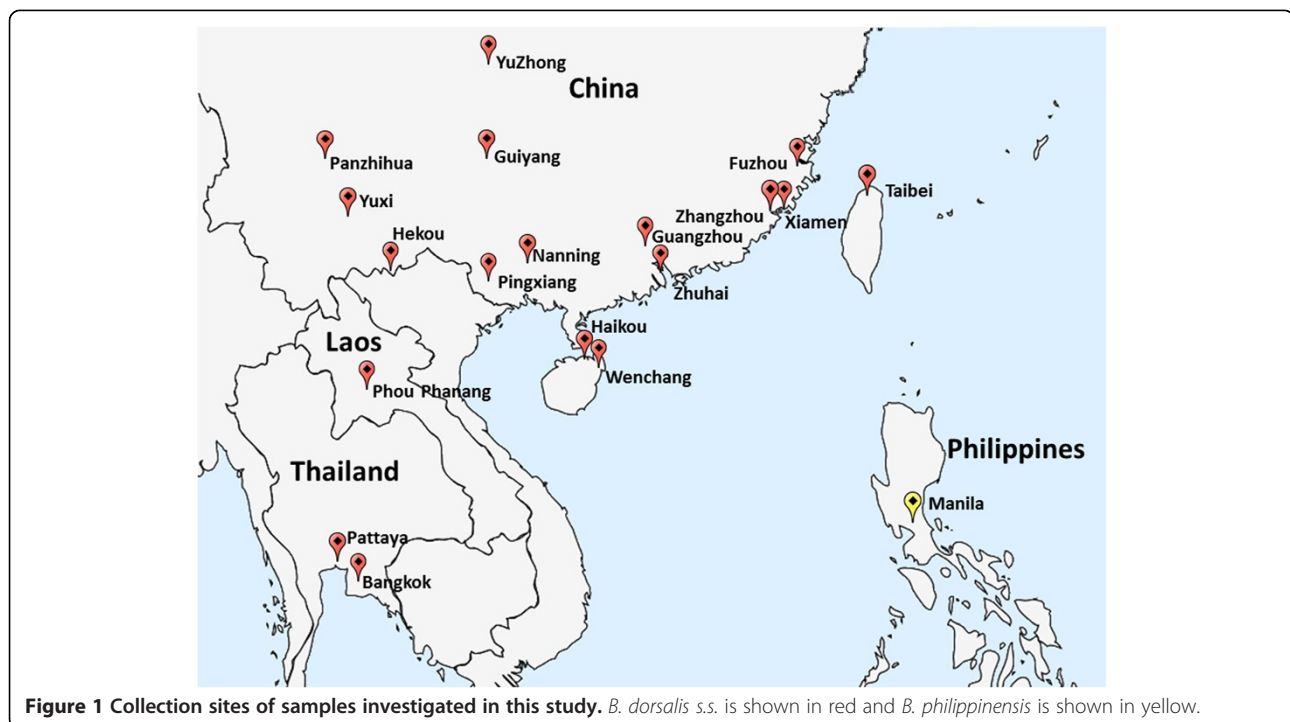
should be attributed to differences in sampling strategies and approaches to analysis. According to Schutze et al., *B. dorsalis s.s.*, *B. papayae*, and *B. philippinensis* should be considered a single species [5]. If this is correct, then analysis of populations sampled from both the ranges of *B. philippinensis* and the *B. dorsalis s.s.* /*B. papayae* transition zone (e.g., Pattaya and Bangkok, in southern Thailand) should provide clues regarding the invasion events in the region [4]. Although mitochondrial DNA markers have been used to examine the changes in the structures of Chinese populations of *B. dorsalis s.s.* over time, the NADH dehydrogenase subunit 1 (*ND1*) markers in *B. dorsalis s.s.* have not yet been used for an evaluation of the distribution area of *B. philippinensis* or the *B. dorsalis s.s.*/*B. papayae* transition zone [15]. In the present investigation, *ND1* was used as a genetic marker to study the genetic structure and origin of the *B. dorsalis s.s.* population within China and the surrounding areas. Mitochondrial DNA is suitable for analyses of population genetics because of its relatively high rate of genetic evolution [16-20]. The study included samples from the *B. philippinensis* distribution area and *B. dorsalis s.s.* /*B. papayae* transition zone to (1) determine the actual number of cryptic species and cryptic lineages of *B. dorsalis s.s.* within China; (2) assess the molecular diversity and genetic architecture of the populations; (3) resolve the controversy of the origin and range expansion of *B. dorsalis s.s.* within China.

## Results

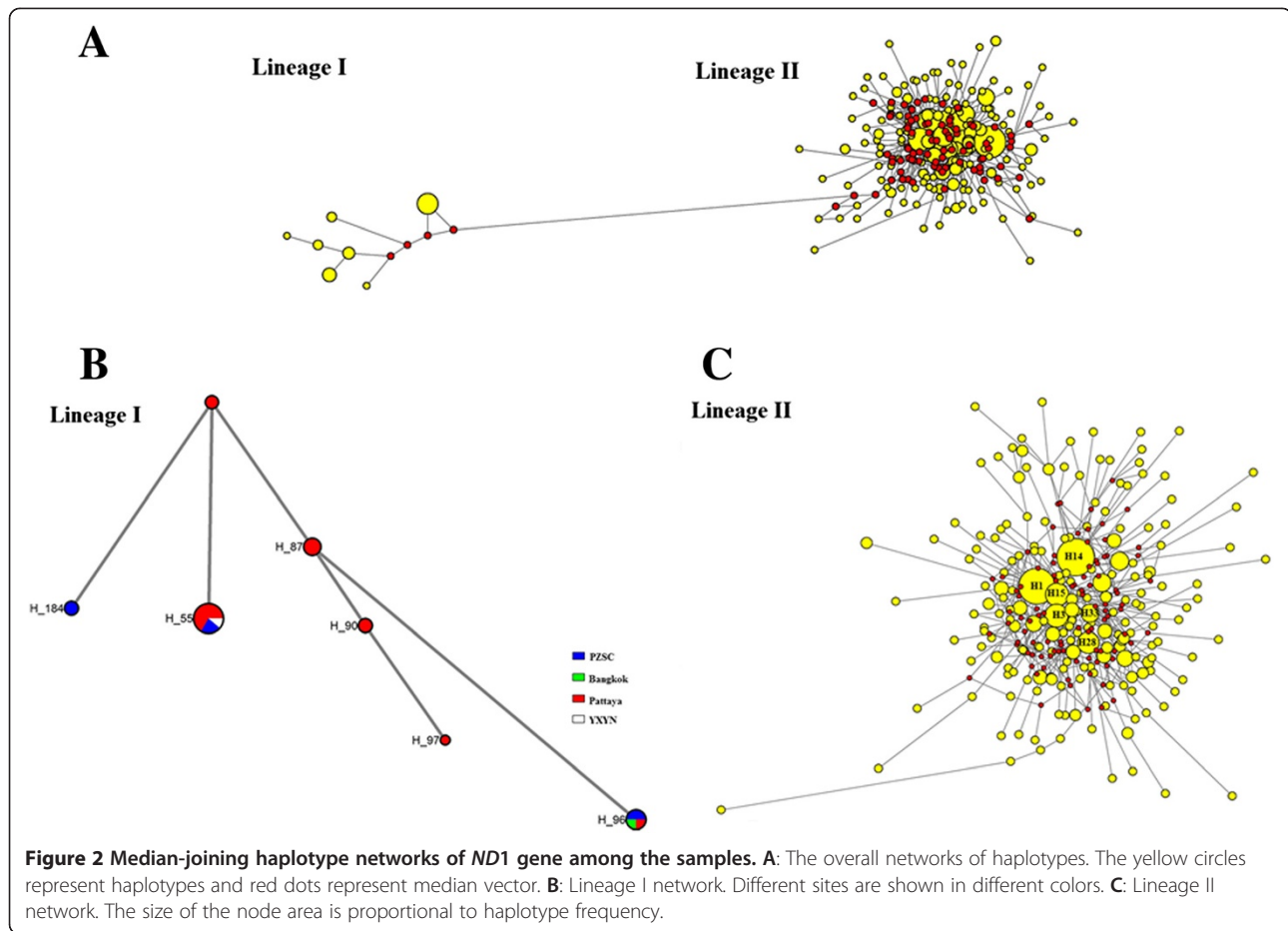
### Two distinct lineages in Chinese *B. dorsalis s.s.*

A total of 19 *B. dorsalis s.s.* populations from China and surrounding areas (Figure 1) were investigated. Analysis of 334 *ND1* sequences from these 19 samples revealed a total of 203 haplotypes, of which 45 were shared and the remaining 158 were single haplotypes. By performing median-joining network analysis, two distinct networks were identified among these 203 haplotypes (Figure 2). The results of the phylogenetic analysis (see Additional file 1: Figure S1) and network analysis clearly showed two distinct lineages, or cryptic species of *B. dorsalis s.s.*, in these samples.

The lineage I network exhibited seven haplotypes (H55, H87, H90, H96, H97, H184, and H186) and the lineage II network showed a star-like pattern with six predominant haplotypes (H1, H3, H14, H15, H28, and H33) (Figure 2). The remaining haplotypes identified from the samples constituted single population groups. The lineage II network mainly included the Pattaya and Panzhuhua populations (Figure 2; see Additional file 2: Table S2). The smaller haplotypes of lineage II network, although connected to the central haplotypes, were found frequently (Figure 2; see Additional file 3: Table S2). These results suggest that most haplotypes have a wide geographic distribution throughout China and the surrounding areas. Rare haplotypes were mainly distributed in Pattaya and Panzhuhua, showing strict geographic distribution characteristics. These findings suggest that the



**Figure 1** Collection sites of samples investigated in this study. *B. dorsalis s.s.* is shown in red and *B. philippinensis* is shown in yellow.



Pattaya and Panzhihua are mix zones of two distinct lineages, which are here treated as putative species.

#### Limited genetic differentiation in geographical populations

To understand the population structure of *B. dorsalis s.s.*, a spatial analysis of molecular variance (SAMOVA) of the *ND1* sequences was performed. The  $F_{CT}$  values in the SAMOVA analysis of the lineage II network suggested the existence of three population groups ( $F_{CT}$  highest for  $K = 3$ ). The 19 populations clustered as follows: Guangzhou, Zhuhai, Nanning, Pingxiang, Yuxi, Hekou, Yuzhong, Guiyang, Haikou, Wenchang, Taibei, Bangkok, and Phou Phanang as Subgroup 1; Fuzhou, Zhangzhou, and Xiamen as Subgroup 2; and Manila alone as Subgroup 3. The group comprising most of mainland China, Hainan island, Taiwan island, and southeastern Asia (including mix zones) was found to be the largest subgroup among all the samples (see Additional file 4: Figure S2).

The AMOVA analysis showed inter-subgroup genetic differentiation to account for 12.31% of all differentiation. The genetic differentiation within populations accounted for 86.38%. The analysis also showed genetic differentiation among the three subgroups to be very limited (1.31%) (Table 1). Differentiation among groups

( $F_{ct}$ ) and within populations ( $F_{st}$ ) was not found to be significant. However, differentiation between groups within the same population ( $F_{sc}$ ) was ( $P < 0.01$ ) (Table 1). Mantel testing indicated significant association between the pairwise coefficient of genetic differentiation ( $F_{st}$ ) and the pairwise geographical distance ( $r = 0.400$ ;  $P = 0.003$ ) within lineage II (see Additional file 2: Table S1). Geographical population cluster levels of genetic differentiation ranged from 0.10207–0.1489 in the  $F_{st}$  values. There was considerable differentiation between the *B. philippinensis* and all other populations ( $F_{st}$  range of 0.10207–0.1489) but moderate amount of differentiation between three other southeast Asian populations (Bangkok, Pattaya, Phou) and all other populations ( $F_{st}$  range of  $-0.03232$ – $0.20702$ ). The amount of differentiation among *B. dorsalis s.s.* populations within China was also moderate ( $F_{st}$  range of  $-0.02021$ – $0.10041$ ) (see Additional file 2: Table S1).

Results obtained from genetic diversity analysis among the 19 geographic populations are summarized in Table 2. The number of haplotypes per population ( $H_p$ ) ranged from 6 to 25. Haplotype diversity ( $H_d$ ) ranged from 0.833 to 1.000, and nucleotide diversity ( $\pi$ ) ranged from 0.00781 to 0.02586. These differences suggested

**Table 1 Analysis of molecular variance (AMOVA) for lineage II**

Source of variation	df	SS	VC	PV	Fixation indices
Among subgroups	2	41.934	0.40244 Va	12.31	$F_{SC} = 0.01491^{**}$
Among populations within subgroups	16	56.715	0.04273 Vb	1.31	$F_{ST} = 0.13620$
Within populations	294	830.102	2.82348 Vc	86.38	$F_{CT} = 0.12312$
Total	312	928.751	3.26865		

df, degrees of freedom; SS, Sum of squares; VC, Variance components; PV, Percentage of variation;  $F_{ST}$ , correlation within populations relative to total.  $^{**}P < 0.01$ .

that all populations retained fairly high levels of genetic diversity. The Manila population showed the most nucleotide diversity ( $\pi$ ), followed by the Pattaya population.

### High migration rates in specific southeast Asian populations

The migration rate of lineage II was estimated in both directions using MIGRATE. Results are shown in Additional file 5: Table S3. The migration rate ranged from 11.9 (from Pingxiang to Guiyang) to 917.1 (from Hekou to Phou). Specific southeast Asian populations (Bangkok, Pattaya, Phou, and Manila) showed very high migration rates. Asymmetric migration was observed between Manila and Pattaya. A similar situation was observed in populations located from Bangkok to China (Guangzhou, Hekou, Fuzhou, Zhangzhou, and Taibei), from Pattaya to China (Guangzhou, Nanning, Hekou, Yuzhong, Xiamen), from

Phou to China (Guangzhou, ZhuHai, Hekou, Yuzhong, Fuzhou, Taibei), and from Manila to China (Nanning, Fuzhou, Zhangzhou). The direction of migration between two sites can be inferred using these asymmetric migration patterns (Figure 3).

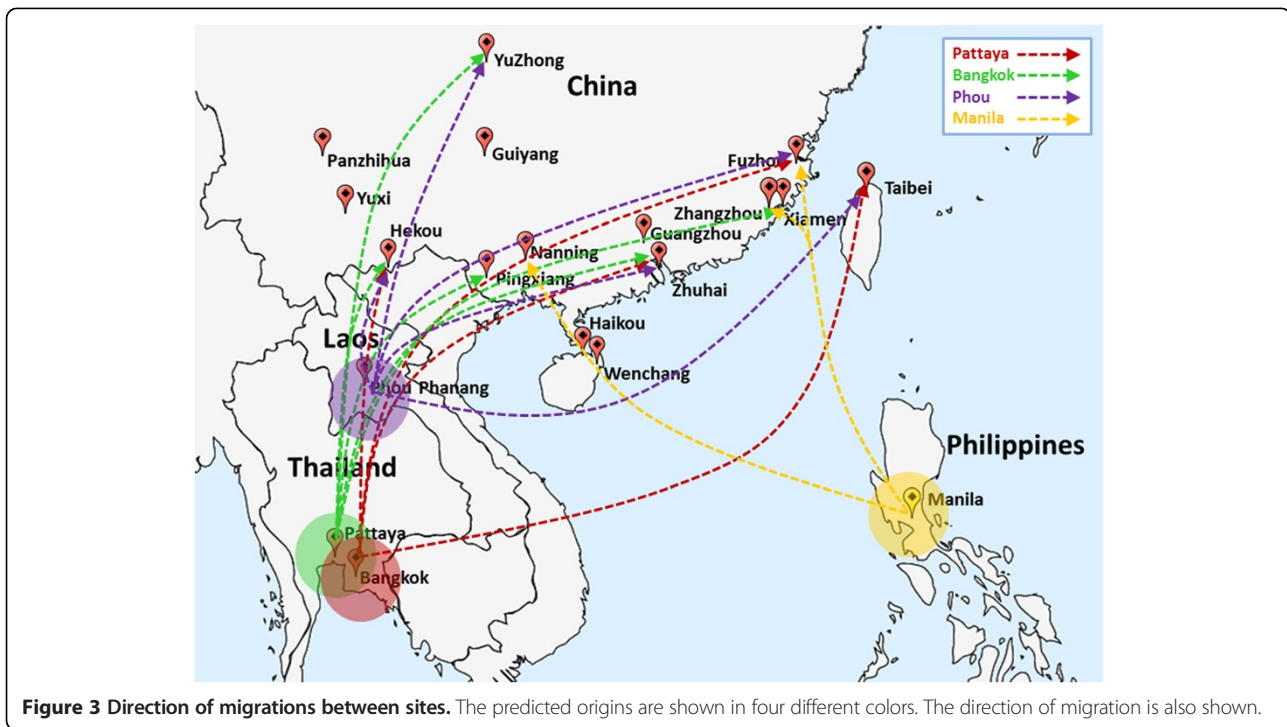
### Marked historical population expansion

The sequence variation of *ND1* genes was used to perform Tajima's *D* test to determine significance of deviations from neutrality. A negative *D* value is produced when more than an expected number of polymorphic sites have low frequencies within a sample. This pattern can be explained by either a recent increase in the size of the population or recent selection [21]. Results indicate negative values ( $D = -1.08524$ ,  $P = 0.17095$ ; Table 3) for lineage II, but these were not found to be statistically significant ( $P > 0.1$ ). This is consistent with a neutral

**Table 2 Genetic diversity parameters of lineage II**

Populations	Code	Numbers	<i>H<sub>p</sub></i>	$\pi \pm SD$	<i>k</i>	<i>H<sub>d</sub></i> $\pm$ SD
China	GZGD	24	18	0.00861 $\pm$ 0.00281	4.341	0.957 $\pm$ 0.031
	ZHGD	15	13	0.01077 $\pm$ 0.00305	5.429	0.971 $\pm$ 0.039
	NNGX	13	13	0.01140 $\pm$ 0.00313	5.744	1.000 $\pm$ 0.030
	PXGX	14	14	0.01180 $\pm$ 0.00336	5.945	1.000 $\pm$ 0.027
	YXYN	18	13	0.01066 $\pm$ 0.00311	5.373	0.928 $\pm$ 0.052
	HKYN	18	17	0.00913 $\pm$ 0.00294	4.601	0.993 $\pm$ 0.021
	YZCQ	12	10	0.01314 $\pm$ 0.00329	6.621	0.955 $\pm$ 0.057
	PZSC	10	10	0.01230 $\pm$ 0.00334	6.200	1.000 $\pm$ 0.045
	GYGZ	19	18	0.01305 $\pm$ 0.00306	6.579	0.994 $\pm$ 0.019
	FZPJ	10	8	0.01067 $\pm$ 0.00298	5.378	0.933 $\pm$ 0.077
	ZZFJ	13	6	0.00804 $\pm$ 0.00212	4.051	0.833 $\pm$ 0.081
	XMFJ	20	11	0.00781 $\pm$ 0.00231	3.937	0.884 $\pm$ 0.054
	HKHN	15	13	0.01270 $\pm$ 0.00305	6.400	0.971 $\pm$ 0.039
	WCHN	22	20	0.01066 $\pm$ 0.00308	5.372	0.991 $\pm$ 0.017
Thailand	TBTW	31	25	0.00821 $\pm$ 0.00267	4.135	0.981 $\pm$ 0.015
	Bangkok	24	22	0.01444 $\pm$ 0.00398	7.275	0.993 $\pm$ 0.014
Laos	Pattaya	8	7	0.02218 $\pm$ 0.00472	11.179	0.964 $\pm$ 0.077
	Phou	19	15	0.00953 $\pm$ 0.00247	4.801	0.977 $\pm$ 0.023
Philippines	Manila	8	8	0.02586 $\pm$ 0.00490	13.036	0.964 $\pm$ 0.077
Total		313	197	0.01181 $\pm$ 0.00379	5.954	0.989 $\pm$ 0.002

$\pi$ , Nucleotide diversity, the average number of nucleotide differences per site between two sequences; *k*, the average number of nucleotide differences; *H<sub>d</sub>*, Haplotype diversity.



**Figure 3** Direction of migrations between sites. The predicted origins are shown in four different colors. The direction of migration is also shown.

mutation hypothesis for the polymorphisms observed in the *ND1* gene. Tests of population expansion using Fu's  $F_s$  were negative and statistically significant across the entire dataset ( $F_s = -13.36645$ ,  $P = 0.01211$ ). The mismatch distribution of lineage II populations was distinctly unimodal (Figure 4), which is strongly indicative of historical population expansion. In lineage I, both tests of population expansion using Fu's  $F_s$  ( $F_s = -24.6437$ ,  $P = 0.93700$ ) and mismatch distribution analysis (Figure 4) suggested that a bottleneck had occurred in lineage I.

The mismatch distribution of lineage II was compatible with the sudden expansion model ( $SSD = 0.020$ ) (Table 3). The parameters of the expansion model were  $\theta_0 = 0.190$ ,  $\theta_1 = 37011.871$ , and  $\tau = 5.701$ . The mismatch distribution of lineage I was not compatible with the expansion model ( $SSD = 0.08150$ ). The parameters of the expansion model were  $\theta_0 = 0.002$ ,  $\theta_1 = 11.597$ , and  $\tau = 9.3203$  (Table 3). The sudden expansion model was rejected for lineage I ( $P < 0.05$ ). The ratio between estimated effective population size after expansion ( $\theta_0$ ) and before expansion ( $\theta_1$ ), which can serve as an estimate of the extent of population growth, was 6589 $\times$  for lineage I and 195,319 $\times$  for lineage II.

## Discussion

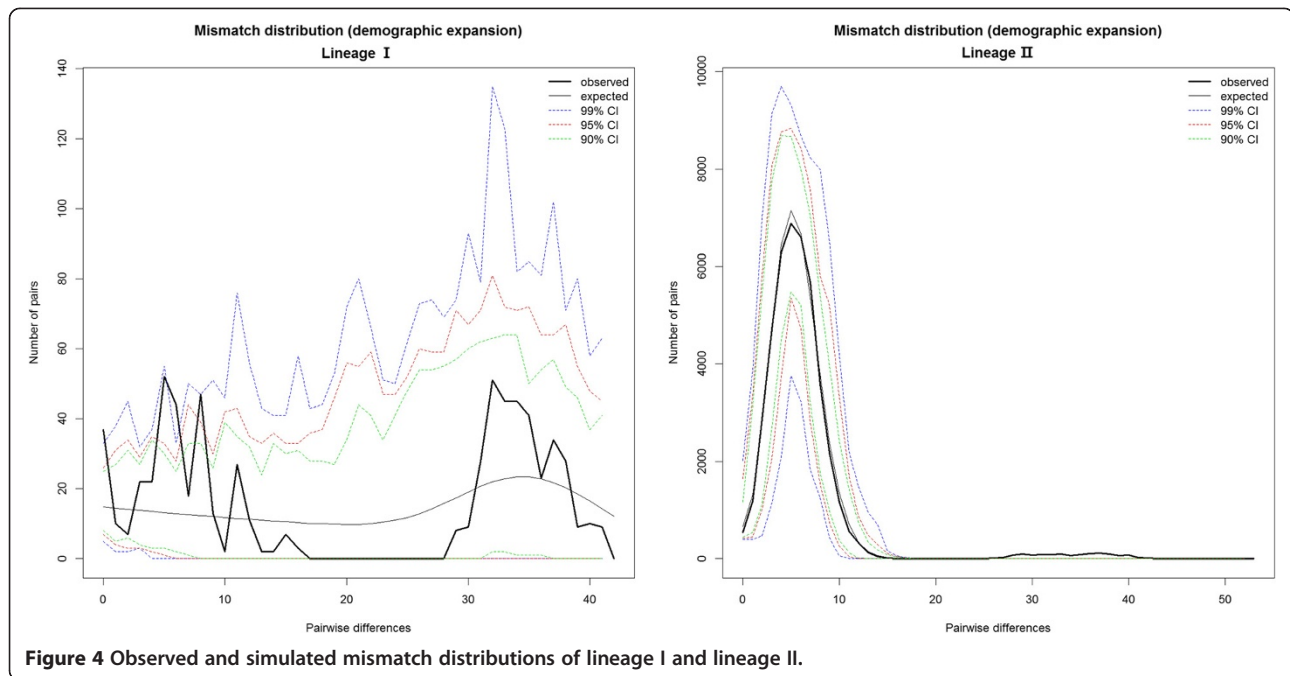
### Molecular markers and sampling strategy

Molecular markers are useful tools for genetic analysis. They can be used to elucidate the mechanisms underlying biological invasion. They are applicable to the identification of cryptic invasions, reconstruction of invasion history, and assessment of the genetic structure of a population [22]. Mitochondrial genes such as the cytochrome oxidase genes (e.g., *COI* and *COII*) have served as molecular markers in studies of the population genetics of invasive insects [23-26]. Although the use of the NADH dehydrogenase gene in such studies is relatively rare, the results of the present study suggest that the NADH dehydrogenase gene is very useful because it shows greater nucleotide substitution rates and more variation than the cytochrome oxidase gene [17-20,27]. Because of this higher rate of variation, the NADH dehydrogenase gene provided a higher resolution in deciphering the geographical population genetics of the recent invasion. NADH dehydrogenase genes have been successfully applied to phylogenetic and population genetic studies of insects, both as an alternative and as a complement to cytochrome oxidase genes [28-30]. Recent investigations

**Table 3** Estimation of sudden population

Group	$\theta_0$	$\theta_1$	$\tau$	$D$	$F_s$	$SSD$
Lineage I	0.00176	11.59668	9.32031	-2.2243	-24.6437	0.08150
Lineage II	0.18949	37011.87112	5.70117	-1.08524	-13.36645	0.01940

$\theta_0$ , effective population size before expansion;  $\theta_1$ , effective population size after expansion;  $\tau$ , 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; \* $P < 0.05$ ; \*\* $P < 0.01$ .



**Figure 4** Observed and simulated mismatch distributions of lineage I and lineage II.

of diversity within the *B. dorsalis s.s.* population have focused primarily on mitochondrial genes encoding subunits I or II of cytochrome oxidase (*CO I* or *CO II*) [14,15,31-33]. It is likely that the *B. dorsalis s.s.* found in China is a recent invasion [1,6,7]. NADH dehydrogenase genes could be better suited to the study of recent invasions. Miller et al. suggested that the gene encoding NADH dehydrogenase subunit 1 (*ND1*) may be suitable for the study of geographical populations and genetic evolution in insects [34]. Based on these facts, *ND1* was chosen for the present study of *B. dorsalis s.s.* population genetics. Earlier claims, in terms of sampling strategy, were not based on an extensive coverage of the distribution of known species. Until recently, *B. dorsalis s.s.*, *B. papayae*, and *B. philippinensis* were thought to belong to different species [2]. However, integrative taxonomy has shown that they are actually a single species [4,5]. For this reason, a larger number of samples were used than in previous studies, and these samples were taken from sites across a broader geographical range (e.g. *B. philippinensis* distribution area and *B. dorsalis s.s./B. papayae* transition zone).

#### Genetic structure, demographic history, and identification of cryptic invasions

Evidence from the *ND1* gene indicated the existence of two distinct lineages that are sympatric within some parts of Pattaya and Panzihua (Figure 2; see Additional file 1: Figure S1). Of the 334 individuals, 21 were found to belong to the lineage I clade and 313 to lineage II. This disequilibrium can be explained by the existence of

two putative species. The first dispersed across a wide geographic area, and the second underwent genetic hitchhiking using external factors (such as infection of maternally inherited microorganisms, such as Wolbachia). The role of Wolbachia in population differentiation and gene flow is known [35]. Previous studies have also found similar phenomena in other species, including Dipteran *Drosophila melanogaster* [36], *Drosophila simulans* [37], *Rhagoletis cerasi* [38], mosquitoes [39], Hemipteran *Bemisia tabaci* [40,41], and Lepidopteran *Danaus chrysippus* (L.) [42,43]. These phenomena suggest that infection of maternally inherited microorganisms may influence random mating among specific taxa, leading to divergence [39,42,43]. If Wolbachia had a role in population structure observed in the current study has yet to be examined. In lineage II, a weak genetic structure was observed due to low differentiation among three groups (Fst value was not significant; see Table 1). This is consistent with the results by Wan et al. [15].

Median-joining networks showed that each clade contains both widespread and localized haplotypes (Figure 2). These patterns suggest recent increases in population size and geographical range. Mismatch distribution analysis was conducted to evaluate the history of species expansion [44]. In order to confirm whether or not the *B. dorsalis s.s.* population had experienced a large-scale population expansion, the ARLEQUIN 3.5 software package was used to analyze mismatch distribution between the haplotypes' base variance. The results showed significant differences between observed and expected values of the haplotypes' base variance along with distinct unimodality distribution

patterns of variation (Figure 4). The estimated effective population size after expansion ( $\theta_1$ ) was 190,000 times higher than before expansion ( $\theta_0$ ) (Table 3). It was here inferred that the *B. dorsalis s.s.* populations of lineage II has experienced rapid expansion in recent years. The average number of nucleotide differences is significantly affected by a population bottleneck, but the number of segregation sites is not [21]. As indicated by the average number of differences in nucleotide sequence ( $k$ ) (Table 2), Chinese *B. dorsalis s.s.* populations showed stronger indications of bottleneck and founder events than the southeast Asian populations (Manila, Pattaya, and Bangkok).

Previous taxonomic analyses of the *B. dorsalis* complex were based on geographical subdivisions due to the difficulties of discrimination using morphological traits [2,45]. According to Drew and Hancock, *B. philippinensis* and *B. occipitalis* but not *B. dorsalis s.s.* are distributed in the Philippines [2]. However, in the present study, haplotype networks and phylogenetic evidence indicated no strong support for geographical subdivision within the *B. philippinensis* population. In this population, three *B. philippinensis* individuals and the entire *B. dorsalis s.s.* population (Guiyang and Xiamen) were found to belong to the same clade. However, five *B. philippinensis* individuals deviated from the *B. dorsalis s.s.* populations (see Additional file 1: Figure S1). Mantel testing indicated that genetic differentiation was significantly influenced by geographical factors ( $r = 0.400$ ;  $P = 0.003$ ). Geographic distance is probably responsible for this partitioning of genetic variation in the Philippines.

### Invasion history

Nucleotide diversity is the ideal index used to evaluate the degree of variation of nucleotide sites between populations [46]. Dependence can be measured using sample size and sequence length, but genetic diversity can be measured using nucleotide diversity. Nucleotide diversity is the average number of differences in the nucleotide sequence among individuals of a given population, compared pairwise. Generally, ancestors show significantly more genetic diversity than derivative populations because of the founder effect [47,48]. According to this principle, it was here inferred that the populations of *B. dorsalis s.s.* from China might have originated in southeast Asia (Manila, Pattaya, and Bangkok) (Table 2). The MIGRATE analysis showed that gene flow occurred from Manila to Bangkok to Phou to China. However, it occurred in an asymmetric manner. The asymmetric migration of gene flow probably indicates colonization events in China that may have involved multiple sources and sites of invasion. Phylogenetic analysis provided another line of evidence for this: numerous southeast Asian individuals belonged to the same clade as Chinese populations (see Additional file 1: Figure S1). This conclusion is

supported by studies showing that southeast Asia might be the region of origin of *B. dorsalis s.s.* [11,49]. As in previous studies, the maximum genetic diversity was observed in southeast Asia and Yunnan Province [11,33,49]. However, symmetric migration patterns were observed at a number of sites, distributed mainly in Yunnan, Guangdong, Fujian, and Taiwan. It is likely that these places may have been the first to be invaded. In conclusion, two distinct lineages of *B. dorsalis s.s.* were identified in the present investigation based on sequence data of *ND1* gene. The asymmetric migration of gene flow indicated multiple invasions and multiple origins.

### Conclusions

Using mitochondrial DNA *ND1* markers, the genetic structure, origin, and invasion history of *B. dorsalis s.s.* were investigated. It was observed that distinct lineages (both minor and major) originated from specific southeast Asian populations. Interestingly, minor lineages have not spread in China. Evidence was found indicating symmetrical migration from southeast Asia to China. Understanding origin and genetic structure of *B. dorsalis s.s.* will possibly assist in the development of effective management strategies to prevent biological invasion. Source-tracking and minor distinct lineage “encounter” approaches may also provide better clues to the design of appropriate control methods, such as introducing natural enemies, to minimize biological invasion of *B. dorsalis s.s.* in China.

### Methods

#### Sample collection, DNA extraction, and sequencing

Adult male flies were collected from 19 natural populations from different locations within China and surrounding areas. All the collections were done during 2007 and 2008. Flies were captured using traps baited with parakairomone methyl eugenol (Table 4; Figure 1). No specific permits were required for these field studies. No specific permission was required to perform these activities in these locations. The locations were not privately owned or protected in any way. The field studies did not involve endangered or protected species. Specimens were preserved in 95% ethanol at  $-20^{\circ}\text{C}$  before use. Flies were identified using morphological identification.

Total DNA was extracted from 334 individual specimens using the Universal Genomic DNA Extraction Kit (TaKaRa). The *ND1* gene fragment of the mitochondrial genome was amplified from all the individuals using the primer *ND1-F* (5'-TTAGTTGCTTGGTTGTGTATTCC-3') and *ND1-R* (5'-GAAAAAGGTAAAAACTCTTTC AAGC-3') [50]. PCR was performed with an initial denaturation step of  $95^{\circ}\text{C}$  for 5 min. The cycle conditions were  $95^{\circ}\text{C}$  for 1 min,  $56^{\circ}\text{C}$  for 1 min, and  $72^{\circ}\text{C}$  for 1.5 min for a total of 35 cycles. PCR ended with a final extension

**Table 4 Sample information used in this study**

Location (province)	No.	GenBank accession no.	Code	Coordinates (lat./long.)	Collect date
Guangzhou (Guangdong)	24	KC413034-KC413057	GZGD	23.13 N 113.23 E	2008.08
Zhuhai (Guangdong)	15	KC413058-KC413072	ZHGD	22.05 N 113.87 E	2008.08
Nanning (Guangxi)	13	KC413073-KC413085	NNGX	22.82 N 108.37 E	2007.08
Pingxiang (Guangxi)	14	KC413086-KC413099	PXGX	22.12 N 106.73 E	2007.08
Yuxi (Yunnan)	19	KC413100-KC413118	YXYN	24.37 N 102.53 E	2007.08
Hekou (Yunnan)	18	KC413119-KC413136	HKYN	22.58 N 103.33 E	2007.08
YuZhong (Chongqing)	12	KC413137-KC413148	YZCQ	29.58 N 106.55 E	2008.08
Panzhihua (Sichuan)	23	KC413149-KC413171	PZSC	26.58 N 101.72 E	2007.08
Guiyang (Guizhou)	19	KC413172-KC413190	GYGZ	26.58 N 106.70 E	2007.08
Fuzhou (Fujian)	10	KC413190-KC413200	FZFI	26.13 N 119.05 E	2007.08
Zhangzhou (Fujian)	13	KC413201-KC413213	ZZFI	24.87 N 117.58 E	2007.08
Xiamen (Fujian)	20	KC413214-KC413233	XMFJ	24.45 N 118.10 E	2007.08
Haikou (Hainan)	15	KC413234-KC413248	HKHN	20.03 N 110.33 E	2008.08
Wenchang (Hainan)	22	KC413249-KC413270	WCHN	19.54 N 110.80 E	2008.08
Taipei (Taiwan)	31	KC413271-KC413301	TBTW	25.05 N 121.50 E	2008.08
Bangkok (Thailand)	25	KC413302-KC413326	Bangkok	13.63 N 101.40 E	2007.08
Pattaya(Thailand)	14	KC413327-KC413340	Pattaya	18.47 N 100.62 E	2007.08
Phou Phanang (Laos)	19	KC413341-KC413359	Phou	18.23 N 102.40 E	2008.08
Manila (Philippines)	8	KC413360-KC413367	Manila	12.87 N 121.77 E	2008.12

of 72°C for 8 min. PCR products were purified and sequenced using PCR primers from Invitrogen Biotechnology Co. (Shanghai, China). Sequences were deposited in GenBank under accession numbers KC413034–KC413367 (Table 4).

#### Data analysis

DNA sequences were assembled using SeqVerter™ and aligned using CLUSTALX 2.0 [51]. They were adjusted manually. Haplotype diversity ( $Hd$ ), nucleotide diversity ( $\pi$ ), average number of differences in nucleotide sequence between haplotypes ( $k$ ), and standard deviation were determined using DNASP 5.0 [52]. To depict the phylogenetic and geographical relationships of the haplotypic sequences, a haplotype network was created with the median-joining method in NETWORK 4.6 using “connection cost” as the parameter criterion (Frequency > 1 criterion, external rooting, and square option was inactive) [53,54]. A phylogenetic tree was constructed using the neighbor joining method implemented in MEGA 5.0 [55].

The coalescence-based program MIGRATE-N 3.5.1 was used to test for and estimate gene flow between populations [56]. MIGRATE-N estimated past migration rates among populations using a matrix model of asymmetric migration rates. MIGRATE-N was used to estimate the immigration rate  $M$  ( $M = m/\mu$ , where  $m$  is the immigration rate per generation and  $\mu$  is the mutation

rate per generation per locus) among populations. One long chain of 100,000,000 generations was set, with the initial 10,000 excluded as burn-in.

Spatial analysis of molecular variance was performed using SAMOVA 1.0 to identify population groups [57]. Sequence datasets, longitude information, and latitude information served as input data. The greatest number of supported 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 ARLEQUIN 3.5 to partition total variance into its components among groups, among populations, and within populations [58]. This was based on the identities of the groups inferred by the SAMOVA analysis. Each of the groupings of populations examined using the AMOVA models explained a significant portion of the molecular variance. Genetic differentiation was estimated by calculating  $F_{st}$  between pairs of populations using ARLEQUIN 3.5.

The correlation of genetic differentiation ( $F_{st}$ ) over geographic distances for all pairs of populations was tested using the Mantel permutation procedure as implemented in ARLEQUIN 3.5 (1000 permutations, significance level  $P < 0.01$ ) [59].

Demographic history was examined using neutrality statistics of Tajima’s  $D$  test, Fu’s  $F_s$  test, and mismatch distribution analysis implemented in ARLEQUIN 3.5



[21,60]. Tajima's *D* and Fu's *F<sub>s</sub>* tests were carried out to examine the deviations from neutrality expected with population expansion. Significant negative *D* and *F<sub>s</sub>* values were considered characteristic of population expansion. Mismatch distribution analysis was performed to evaluate the frequency distribution of pairwise differences between sequences. A unimodal approximate Poisson-like distribution is expected for populations that have experienced demographic expansion in the recent past, but multimodal frequency distribution is expected for populations at equilibrium [61]. Population expansion was further assessed by examining the fitness between the observed and expected frequency distribution with the Harpending's raggedness index and sum of squared difference (*SSD*) statistics using ARLEQUIN 3.5 [62,63].

## Additional files

**Additional file 1: Figure S1.** Subgroups of lineage II. The three subgroups are shown by different colors.

**Additional file 2: Table S1.** Haplotypes of *ND1* identified in each population.

**Additional file 3: Table S2.** Coefficient of genetic differentiation (under diagonal) and geographic distances (above diagonal) of lineage II.

**Additional file 4: Figure S2.** Neighbor-joining tree of lineage I and II.

**Additional file 5: Table S3.** Immigration rate between population pairs estimated using MIGRATE.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

JTL conceived the ideas for the study; SYB, JM, HLH, XFL, and JTL collected the samples; ZZW and HML performed the experiments; ZZW analyzed the data and composed the manuscript. All authors have read and approved the final manuscript.

## Acknowledgements

We are indebted to Dr. Xiao-Wen Cheng of Miami University of Ohio for advice and helpful review of the manuscript. This work was funded by the National Natural Science Foundation of China under Grant 30671378.

## Author details

<sup>1</sup>Institute for Management of Invasive Alien Species, 314 Yingdong teaching building, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, PR China. <sup>2</sup>College of Life Sciences, Zhongkai University of Agriculture and Engineering, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, PR China. <sup>3</sup>Guangdong Inspection and Quarantine Technology Center, Guangdong Entry-Exit Inspection and Quarantine Bureau, Tower A, Guojian Building, No.66 Huacheng Avenue, Zhujiang Xincheng, Guangzhou 510225, PR China.

Received: 26 October 2013 Accepted: 14 March 2014

Published: 21 March 2014

## References

- White IM, Elson-Harris MM: *Fruit flies of economic significance: their identification and bionomics*. Wallingford UK: CAB International; 1992.
- Drew RAI, Hancock DL: The *Bactrocera dorsalis* complex of fruit flies (Diptera: tephritidae: dacinae) in Asia. *Bull Entomol Res Suppl* 1994, **2**(i-iii):1-68.
- Iwahashi O: Aedeagal length of the oriental fruit fly, *Bactrocera dorsalis* (Hendel) (Diptera: tephritidae), and its sympatric species in Thailand and

- the evolution of a longer and shorter aedeagus in the parapatric species of *B. Dorsalis*. *Appl Entomol Zool* 2001, **36**(3):289-297.
- Krosch MN, Schutze MK, Armstrong KF, Boontop Y, Boykin LM, Chapman TA, Englezou A, Cameron SL, Clarke AR: Piecing together an integrative taxonomic puzzle: microsatellite, wing shape and aedeagus length analyses of *Bactrocera dorsalis* s.l. (Diptera: tephritidae) find no evidence of multiple lineages in a proposed contact zone along the Thai/Malay Peninsula. *Syst Entomol* 2013, **38**(1):2-13.
- Schutze MK, Krosch MN, Armstrong KF, Chapman TA, Englezou A, Chomic A, Cameron SL, Hailstones D, Clarke AR: Population structure of *Bactrocera dorsalis* s.s., *B. papayae* and *B. philippinensis* (Diptera: tephritidae) in southeast Asia: evidence for a single species hypothesis using mitochondrial DNA and wing-shape data. *BMC Evol Biol* 2012, **12**:130.
- DE H: *Pacific insects monograph 31: The fruit flies (Diptera: tephritidae) of Thailand and bordering countries*. Honolulu: Bernice P Bishop Museum; 1973.
- Xie YZ: Study on the trypetidae or fruit-flies of China. *Sinensia* 1937, **2**:103-226.
- Wang XJ: The fruit flies of the East Asian region. *Acta Zootaxon Sinica* 1996, **54**:49-54.
- Ye H, Chen P: Expansion tendency of *Bactrocera dorsalis* in China. In *Research on biological invasions in China*. Edited by Wang FH, Guo JY, Zhang F. Beijing: Science Press; 2009:90-92.
- Xie Q, Zhang RJ: Study advance on biology and ecology of *Bactrocera dorsalis* (Hendel) and its control. *Ecol Sci* 2005, **24**:52-56.
- Aketarawong N, Bonizzoni M, Thanaphum S, Gomulski LM, Gasperi G, Malacrida AR, Gugliemino CR: Inferences on the population structure and colonization process of the invasive oriental fruit fly, *Bactrocera dorsalis* (Hendel). *Mol Ecol* 2007, **16**(17):3522-3532.
- Li Y, Li Z, Wu G, Wang H, Deng Y, Gong X: A primary on the population genetics relationship of *Bactrocera dorsalis* (Diptera: tephritidae) based on *mtDNA ND 6 gene*. *Biol Control*: Chinese J; 2009:42-50.
- Li W, Yang L, Tang K, Zeng L, Liang G: Microsatellite polymorphism of *Bactrocera dorsalis* (Hendel) populations in China. *Acta Entomol Sin* 2007, **50**(12):1255-1262.
- Li Y, Wu Y, Chen H, Wu J, Li Z: Population structure and colonization of *Bactrocera dorsalis* (Diptera: tephritidae) in China, inferred from *mtDNA COI* sequences. *J Appl Entomol* 2012, **136**(4):241-251.
- Wan X, Nardi F, Zhang B, Liu Y: The oriental fruit Fly, *Bactrocera dorsalis*, in China: origin and gradual inland range expansion associated with population growth. *PLoS One* 2011, **6**(10):e25238.
- Behura SK: Molecular marker systems in insects: current trends and future avenues. *Mol Ecol* 2006, **15**(11):3087-3113.
- Cameron SL, Downton M, Castro LR, Ruberu K, Whiting MF, Austin AD, Diement K, Stevens J: Mitochondrial genome organization and phylogeny of two vespid wasps. *Genome* 2008, **51**(10):800-808.
- Hua J, Li M, Dong P, Cui Y, Xie Q, Bu W: Comparative and phylogenomic studies on the mitochondrial genomes of pentatomomorpha (Insecta: hemiptera: heteroptera). *BMC Genomics* 2008, **9**:610.
- Nardi F, Carapelli A, Dallai R, Frati F: The mitochondrial genome of the olive fly *Bactrocera oleae*: two haplotypes from distant geographical locations. *Insect Mol Biol* 2003, **12**(6):605-611.
- Negrisola E, Babbucci M, Patarnello T: The mitochondrial genome of the ascalaphid owlfly *Libelloides macaronius* and comparative evolutionary mitochondrialomics of neuropterid insects. *BMC Genomics* 2011, **12**:221.
- Tajima F: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989, **123**(3):585-595.
- Miura O: Molecular genetic approaches to elucidate the ecological and evolutionary issues associated with biological invasions. *Ecol Res* 2007, **22**(6):876-883.
- Grapputo A, Boman S, Lindstrom L, Lyytinen A, Mappes J: The voyage of an invasive species across continents: genetic diversity of North American and European Colorado potato beetle populations. *Mol Ecol* 2005, **14**(14):4207-4219.
- Havill NP, Montgomery ME, Yu GY, Shiyake S, Caccione A: Mitochondrial DNA from hemlock woolly adelgid (Hemiptera: adelgidae) suggests cryptic speciation and pinpoints the source of the introduction to eastern north America. *Ann Entomol Soc Am* 2006, **99**(2):195-203.
- Hu J, De Barro P, Zhao H, Wang J, Nardi F, Liu S: An extensive field survey combined with a phylogenetic analysis reveals rapid and widespread invasion of Two alien whiteflies in China. *PLoS One* 2011, **6**(1):e16061.
- Ueda S, Brown JK: First report of the Q biotype of *Bemisia tabaci* in Japan by mitochondrial cytochrome oxidase I sequence analysis. *Phytoparasitica* 2006, **34**(4):405-411.

27. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P: Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann Entomol Soc Am* 1994, **87**(6):651–701.
28. Birungi J, Munstermann LE: Genetic structure of *Aedes albopictus* (Diptera: culicidae) populations based on mitochondrial ND5 sequences: evidence for an independent invasion into Brazil and United States. *Ann Entomol Soc Am* 2002, **95**(1):125–132.
29. Elfekih S, Makni M, Haymer DS: Genetic diversity of ND5 mitochondrial patterns in *Ceratitis capitata* (Diptera: tephritidae) populations from Tunisia. *Ann Soc Entomol Fr* 2010, **46**(3–4):464–470.
30. Rach J, DeSalle R, Sarkar IN, Schierwater B, Hadrys H: Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *P Roy Soc B-Biol Sci* 2008, **275**(1632):237–247.
31. Chen P, Ye H: Relationship among five populations of *Bactrocera dorsalis* based on mitochondrial DNA sequences in western Yunnan. *China. J Appl Entomol* 2008, **132**(7):530–537.
32. Liu J, Shi W, Ye H: Population genetics analysis of the origin of the oriental fruit fly, *Bactrocera dorsalis* Hendel (Diptera: tephritidae), in northern Yunnan province. *China. Entomol Sci* 2007, **10**(1):11–19.
33. Shi W, Kerdelhue C, Ye H: Population genetics of the oriental fruit fly, *Bactrocera dorsalis* (Diptera: tephritidae), in Yunnan (China) based on mitochondrial DNA sequences. *Environ Entomol* 2005, **34**(4):977–983.
34. Miller NJ, Dorhout DL, Rice ME, Sappington TW: Mitochondrial DNA variation and range expansion in Western Bean Cutworm (Lepidoptera: noctuidae): no evidence for a recent population bottleneck. *Environ Entomol* 2009, **38**(1):274–280.
35. Sun X, Cui L, Li Z: Diversity and phylogeny of *Wolbachia* infecting *Bactrocera dorsalis* (Diptera: tephritidae) populations from China. *Environ Entomol* 2007, **36**(5):1283–1289.
36. Hoffmann AA, Clancy DJ, Merton E: Cytoplasmic incompatibility in Australian populations of *Drosophila melanogaster*. *Genetics* 1994, **136**(3):993–999.
37. Montchamp-Moreau C, Ferriere JF, Jacques M: Geographic distribution and inheritance of three cytoplasmic incompatibility types in *Drosophila simulans*. *Genetics* 1991, **129**(2):399–407.
38. Riegler M, Stauffer C: *Wolbachia* infections and superinfections in cytoplasmically incompatible populations of the European cherry fruit fly *Rhagoletis cerasi* (Diptera, tephritidae). *Mol Ecol* 2002, **11**(11):2425–2434.
39. Kittayapong P, Baisley KJ, Baimai V, O'Neill SL: Distribution and diversity of *Wolbachia* infections in Southeast Asian mosquitoes (Diptera: culicidae). *J Med Entomol* 2000, **37**(3):340–345.
40. De Barro PJ: Genetic structure of the whitefly *Bemisia tabaci* in the Asia-Pacific region revealed using microsatellite markers. *Mol Ecol* 2005, **14**(12):3695–3718.
41. Ahmed MZ, Ren S, Mandour NS, Greeff JM, Qiu B: Prevalence of *wolbachia* supergroups A and B in *Bemisia tabaci* (Hemiptera: aleyrodidae) and some of its natural enemies. *J Econ Entomol* 2010, **103**(5):1848–1859.
42. Lushai G, Allen JA, Goulson D, MacLean N, Smith DAS: The butterfly *Danaus chrysippus* (L.) in east Africa comprises polyphyletic, sympatric lineages that are, despite behavioural isolation, driven to hybridization by female-biased sex ratios. *Biol J Linn Soc* 2005, **86**(1):117–131.
43. Smith DAS, Gordon IJ, Allen JA: Reinforcement in hybrids among once isolated semispecies of *Danaus chrysippus* (L.) and evidence for sex chromosome evolution. *Ecol Entomol* 2010, **35**(1):77–89.
44. Rogers AR: Genetic evidence for a pleistocene population explosion. *Evolution* 1995, **49**(4):608–615.
45. Clarke AR, Armstrong KF, Carmichael AE, Milne JR, Raghu S, Roderick GK, Yeates DK: Invasive phytophagous pests arising through a recent tropical evolutionary radiation: the *Bactrocera dorsalis* complex of fruit flies. *Annu Rev Entomol* 2005, **50**:293–319.
46. Nei M, Li WH: Mathematical model for studying genetic variation in terms of restriction endonucleases. *P Natl Acad Sci Usa* 1979, **76**(10):5269–5273.
47. Grant WS, Bowen BW: Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *J Hered* 1998, **89**(5):415–426.
48. Templeton AR: The theory of speciation via the founder principle. *Genetics* 1980, **94**(4):1011–1038.
49. Shi W, Kerdelhue C, Ye H: Population genetic structure of the oriental fruit fly, *Bactrocera dorsalis* (Hendel) (Diptera: tephritidae) from Yunnan province (China) and nearby sites across the border. *Genetica* 2010, **138**(3):377–385.
50. Nardi F, Carapelli A, Dallai R, Roderick GK, Frati F: Population structure and colonization history of the olive fly, *Bactrocera oleae* (Diptera, tephritidae). *Mol Ecol* 2005, **14**(9):2729–2738.
51. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: Clustal W and clustal X version 2.0. *Bioinformatics* 2007, **23**(21):2947–2948.
52. Librado P, Rozas J: DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009, **25**(11):1451–1452.
53. Bandelt H, Forster P, Roehl A: Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 1999, **16**(1):37–48.
54. Forster P, Torroni A, Renfrew C, Roehl A: Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol Biol Evol* 2001, **18**(10):1864–1881.
55. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011, **28**(10):2731–2739.
56. Beerli P: Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 2006, **22**(3):341–345.
57. Dupanloup I, Schneider S, Excoffier L: A simulated annealing approach to define the genetic structure of populations. *Mol Ecol* 2002, **11**(12):2571–2581.
58. Excoffier L, Lischer HEL: Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010, **10**(3):564–567.
59. Mantel N: The detection of disease clustering and a generalized regression approach. *Cancer Res* 1967, **27**(2):209–220.
60. Fu Y: Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 1997, **147**(2):915–925.
61. Stopar K, Ramsak A, Trontelj P, Malej A: Lack of genetic structure in the jellyfish *Pelagia noctiluca* (Cnidaria: scyphozoa: semaeostomeae) across European seas. *Mol Phylogenet Evol* 2010, **57**(1):417–428.
62. Harpending HC: Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* 1994, **66**(4):591–600.
63. Rogers AR, Harpending H: Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 1992, **9**(3):552–569.

doi:10.1186/1471-2148-14-55

Cite this article as: Wu et al.: Sequence analysis of mitochondrial ND1 gene can reveal the genetic structure and origin of *Bactrocera dorsalis* s. s.. *BMC Evolutionary Biology* 2014 **14**:55.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

