Genetic variation and population structure of the Oriental fruit fly, *Bactrocera dorsalis* (Diptera: Tephritidae) in China and Southeast Asia

Shi Wei

Bactrocera dorsalis (Hendel) (Diptera: Tephritidae), one species of phytophagous insects in Tephritidae family, is mainly distributed in Asia-Pacific Ocean regions. Most distributional areas of *B. dorsalis* in Asia fall into 24° N and 25° S (low latitude areas), where is typical tropical climate. A stable population stay in these areas. Decreasing distributional areas of the fly are in regions between 24° N and 30° N (high latitude areas), where many populations distribute seasonally. We compared population structure and gene diversity of China (East Asia) and Southeast Asia countries based on microsatellite markers. Our data suggested that populations of low latitude areas of Asia maybe pertain the native range of *B. dorsalis.* Some Southeast Asia *B.dorsalis* populations including Vietnam and Myanmar perhaps belong to the native range of this species. A south to north colonize pathway was proposed for *B. dorsalis* in Asia. In China, populations formed a homogenous unite and an extensive gene flow were existed among China populations. Yunnan-R, southwest populations of China and Yunnan-H, Guangxi, two south China populations show clear features of entrance that the fly was introduced from Southeast Asia.

Introduction:

Phytoghagous flies of the Tephritidae family also namely “true fruit flies”, are most important pests for fruits and vegetables. With more than 4000 species founded, one-third flies of this family can attack soft fruits, including many commercially and economically important ones. *Bactrocera dorsalis*, the oriental fruit flies, is one of the most notorious one in Tephritidae family. This species is a major highly polyohagous pest that attacks more than 250 plants including a number of commercially grown fruits, such as melon, banana, mango and guava (Li and Ye 2000). Because its wide host range and demographic strategy traits towards the kind of the r-k gradient of growth curve, *Bactrocera dorsalis* would be adapted for growth and establishment in near optimal environmental condition, taking over the most productive niches and dispersing very fast (Aketarawong et al. 2007).

- This fly is first recorded in Taiwan (Hardy 1973; Drew and Hancock 1994) in the beginning of last century, and over the following 90 years, it has expanded throughout most regions in Asia covering Central, East and Southeast Asia and also around the Pacific Ocean (see Fig.1) (Hardy 1973). Most distributional countries of this fly in Asia and Pacific Ocean falling into the tropical climate regions (between 24ºN and 25ºN) (Wang 2002) including Indonesia, Philippines, Laos, Cambodia, Vietnam, south of Myanmar, Thailand, Srilanka, Malaysia, Brunei, south of India and Bangladesh, south of Taiwan and China, northern Mariana Island, Nauru and Guam. (IIE 1994 a, b; EPPO, 2003). A small part of distributional areas of the fly falling into regions between 24ºN and 30ºN including central and southwest of China, north of India and Myanmar, Bhutan, Nepal, north of Pakistan and Bangladesh and Hawaii. When the latitude higher than 30° N, the distributional areas of *B. dorsalis* are very few excluding Japan (has been eradication in 1986) and north of Pakistan (IIE 1994 a,b; EPPO, 2003, Hou et al. 2005).
- In China, the first reported appearance of the fly occurred in 1934 in Hainan province (Xie 1937). However, in successive 20 years from 1934, the reporting about this fly in China is absent. After 1950, *B. dorsalis* infestations have been reported in southwest and southeast part of China (Wang 1996). At the present time, the fly has distributed in 8 provinces of mainland in China (including Gunangxi, Guangdong, Fujian, Yunnan, Guizhou, Sichuan, Hunnan and Hainan) and Taiwan (Liang and Chen, 2003).

Fig.1 World wide distribution of *Bactrocera dorsalis*

- Population genetic data are important to infer the origin and understand the colonization success and history of the invasion pest (Virgilio et al.2010). In terms of genetic studies of this fly, more attentions were concerned about species define (Nakahara et al. 2000, 2001; Richard et al. 2008), intraspecific variation of different geographic populations (Shi et al. 2005, 2007, 2010; Liu et al. 2007; Nakahara et al. 2008), but identifying the source origin and understand the colonization pathway for this species remained relatively unexplored (Aketarawong et al. 2007). In the study of Aketarawong, it is hypothesized that China pertain the source area of *B. dorsalis* and suggested an East-West colonization route to other regions o Asia. However, this hypothesis need be confirmed by extensive sampling of the China region, because just one sample site appeared in this study**.**

  Our present study is based on the more widest geographical sampling for the oriental fruit flies, consisting of 16 populations including more sample site in China and several major Southeast Asia countries where the *B. dorsalis* occurs, We applies multilocus microsatellite loci markers which have been used widely to insect populations genetic studying including population structure, phylogeography and invasion biology (Nardi et al. 2005; Roderick 2004). We use these makers to infer the genetic variability in a more large geographic scale and to understand the source areas and colonize pathway for this species continually.

Materials and methods

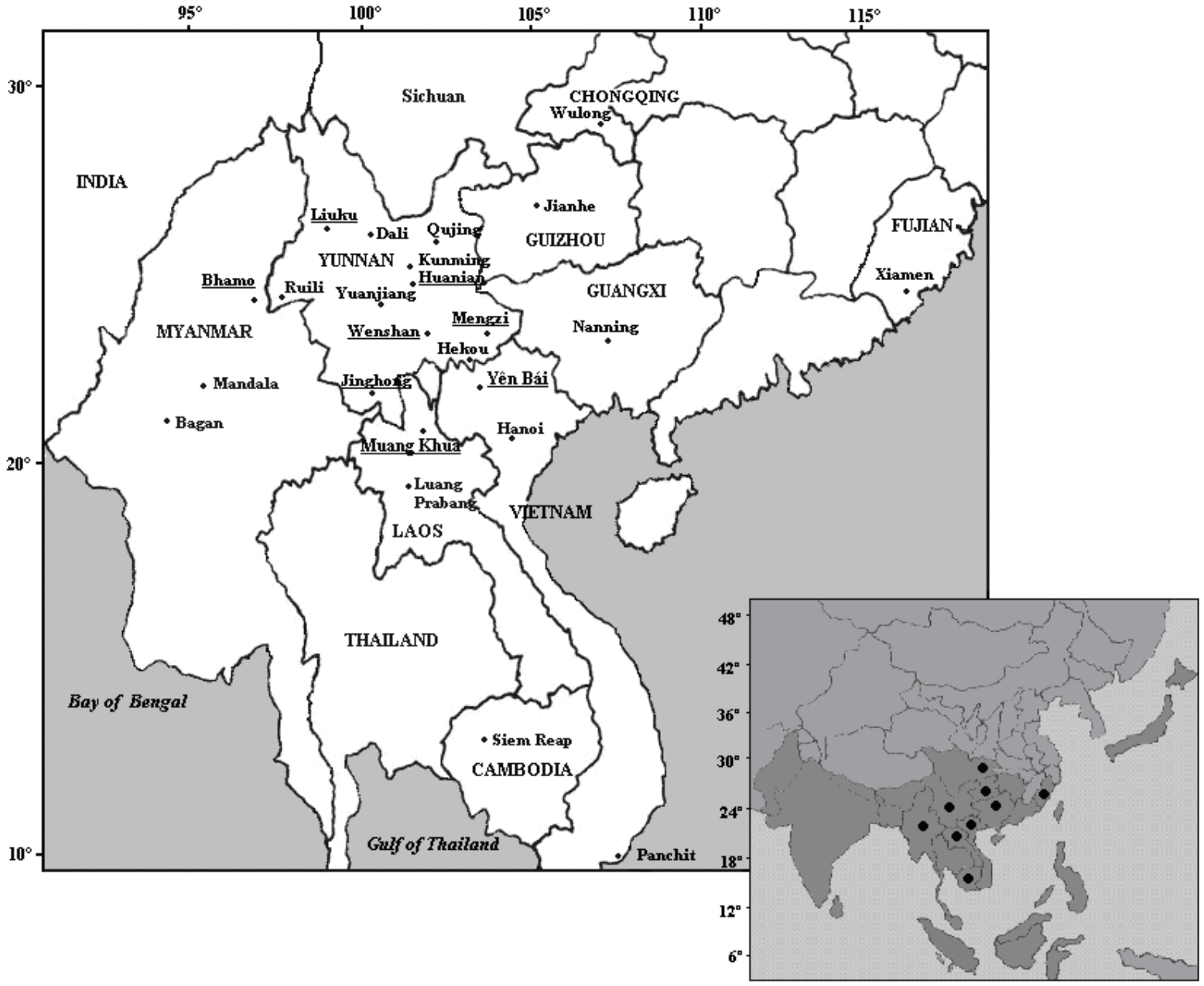
Sample collection

samples of *B. dorsalis* were collected from 16 localities from China and Southeast Asia countries (Fig.2). Ten of the samples were composed of 20 individuals, two of Yunnan samples were 25 individuals, also other two samples of Yunnan were 23 individuals. Two Myanmar samples were 15 individuals. In all sample sites, the latitude are lower than 24° N are all shade in Table 1. These populations hereafter are named "lower latitude populations". But other populations occurring at regions where the latitude higher than 24° N hereafter are named "high latitude populations". All the samples were obtained as ethanol preserved adults and DNA was extracted from each single fly according to the method of Wei Shi et al. (2005).

Microsatellite analysis

Nine microsatellite loci were used in this study. BD47 loci had been outlined in Aketarawong et al. (2006), six loci: MS3, MS3B, MS4, MS5, MS6, MS12A were reported in Dai et al. (2004) and 618A, BO-D48 were described in Li et al. (2007). Primer sequences and the methods used for DNA amplification, electrophoresis and allele scoring using an automated ABI PRISM310 Genetic Analyser were as reported in Aketarawong et al. (2006). An individual was declared null (not amplifying at a locus) only after at least two amplification failures.

- **Fig. 2** (see below) Our sampling sites of *B. dorsalis* numbered according to Table 1, the samples only applied with mitochondrial haplotype marker was underlined. The small figure low right corner is distribution map of *B. dorsalis* in Asia. Areas of dark grey represents that the *B. dorsalis* distributional countries or regions. represents the sample site in this study. Code of each population is Yunnan-Yuanjian (YY), Yunnan-Kunming (YK), Yunnan-Dali (YD), Yunnan-Qujing (YQ), Yunnan-Ruili (YR), Yunnan-Hekou (YH), Guizhou (GZ), Guangxi (GX), Chongqing (CQ), Fujian (FJ), Myanmar-Mandala (MM), Myanmar-Bagan (MB), Vietnam –Hanoi (VH), Vietnam –Panchit (VP), Laos-Luang Prabang (LL), Laos-Muang Khua (LM), Cambodia -Siem Reap (CA).


95°
100°
105°
110°
115°
30°
20°
10°
Sichuan
CHONGQING
Wulong
INDIA
Jianhe
Liuku
Dali
Qujing
GUIZHOU
FUJIAN
YUNNAN
Kunming
Huanian
Xiamen
Bhamo
Ruili
Yuanjiang
GUANGXI
Mengzi
MYANMAR
Wenshan
Nanning
Hekou
Mandala
Jinghong
Yên Bái
Bagan
Hanoi
Muang Khua
Luang
Prabang
VIETNAM
LAOS
THAILAND
Bay of Bengal
Siem Reap
CAMBODIA
Gulf of Thailand
Panchit
48°
42°
36°
30°
24°
18°
12°
6°

- For each population, the following genetic diversity estimates were calculated as averages over all loci with Genepop 4.0 (Rousset 2008): mean number of alleles (na); number of private alleles (np), frequency of private alleles (AP); observed hetero zygosity (HO) and expected heterozygosity (HZ). Gene diversity (HS), Allelic richness (RS) and the inbreeding index FIS estimates were measured using FSTAT 2.9.3.2 (Goudet 2001). Nucleotide diversity and within population mean number of pairwise differences were assessed using ARLEQUIN3.11 (Excoffieretal 2005). Genepop 4.0 was also used to test for linkage disequilibrium between pairs of loci in each population and deviation from the Hardy-Weinberg equilibrium, together with their critical levels after the sequential Bonferroni test (Rice 1989). The frequency of null alleles was calculated using the Brook field estimation (Brookfield 1996) in FreeNA (Brookfield 1996). The same program was used to measure the degree of genetic differentiation between pairs of population, in terms of pairwise *FST* values of weir (1996) using ENA correction described in Chapllis and Estoup (2007), the statistical significance of each value was assessed by the comparison of the observed value with the values obtained in 10000 matrix permutations.

The Bayesin approach as implemented in STRUCTURE 2.2 (Pritchard et al. 2000) was used to infer the clustering of 16 populations.

The GENECLASS 2.0 software (Piry et al. 2004) was to calculate population assignment and exclusion test including the calculation of probability of origin for each individual.

The program Bottleneck version 1.2.02 (Cornuet &Luikart 1996) was used in the attempt to recognize in our samples the effect of a recent bottleneck. Two mutation models, considered appropriate for microsatellites (Piry et al. 1999), were applied: the strict Stepwise Mutation Model (SMM) and the Two-phase Model (TPM). For the TPM, a model that includes both 90% single-step mutations and 10% multiple step mutations was used. Significant deviations in observed heterozygosity over all loci were tested using a nonparametric Wilcoxon test.

- ***Genetic diversity***

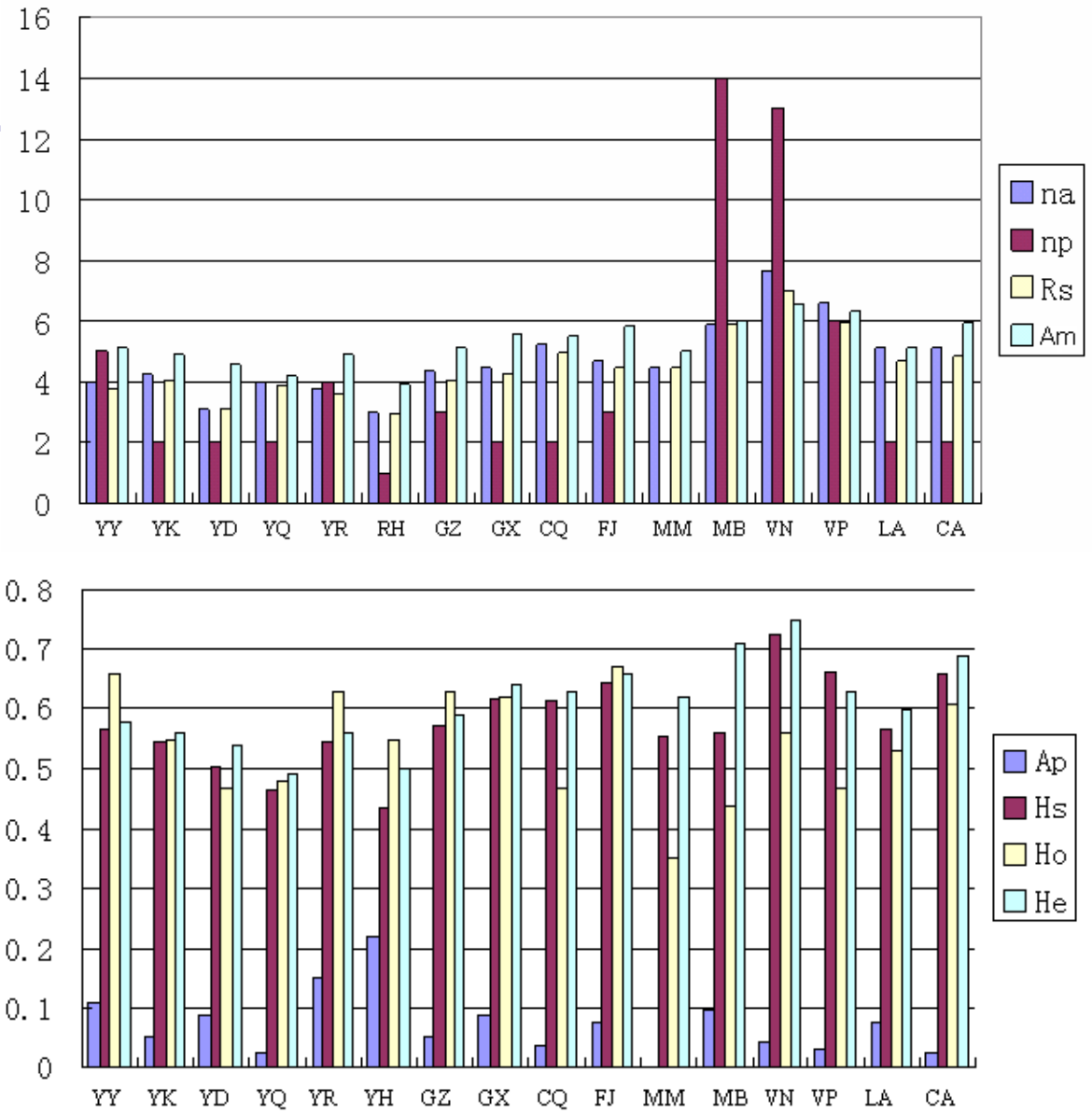
Our data are based on 9 microsatellite loci detected in 326 *B. dorsalis* fly across 16 populations Over all populations, the 9 loci show different levels of polymorphism both in terms of number of alleles (from 2 to 12) and allele size range: 618A: 23 alleles (133-173bp); BD47: 12 alleles (103-128bp) BO-DA8: 18 alleles (148-174bp); MS3: 20 alleles (235-285bp), MS3B: 18 alleles (217-245bp), MS4: 19 alleles (129-173bp), MS5: 17 alleles (119-157bp), MS6: 11 alleles (160-194bp), MS12: 10 alleles (161-189bp).

Fig 3 shows the genetic variability over the 9 loci for each *B. dorsalis* populations. The samples of Hanoi presents the highest genetic variability among 16 populations, with a mean number of alleles of 7.67, 13 private alleles, a highest allelic richness(6.96), genetic diversity (0.725) and average number differences within population (6.54). Followed Hanoi population, Panchit population also shows a very high genetic variability in all 16 populations. Bangan population of Myanmar also processes a high gene variability such as 14 private alleles and 5.89 mean numbers of alleles and so on. Compared with all population of Southeast Asia, we can find Mandala of Myanmar show a very lower genetic variability.

- Among five populations of China, Fujian presents a relative higher genetic variability in terms of number of private alleles (3), gene diversity (HS=0.645), mean observed heterozygosity (0.67), mean expected heterozygosity (0.66) and average number differences within population(5.81). Guangxi populations also have relative higher genetic variability. For six Yunnan populations, Yuanjiang and Ruili population show higher genetic variability, but Hekou population presents the lowest values of genetic variability. Comparing with high latitude populations, low latitude populations processes richer genetic variability, including mean number of alleles, the number of private alleles, allelic richness, Ho, HE and average number differences within population (also see Fig.4, each gene variability index above are all compare by U-test and all $P<0.05$).

Fig.3 Genetic variability estimates in field-collected samples of *B. dorsalis* from different geographical regions based on microsatellite data.

na, mean number of alleles; np, number of private alleles; Ap, mean frequency of private alleles; RS, allelic richness and standard deviation; HS, gene diversity; HO, mean observed heterozygosity; HE, mean expected heterozygosity; Am is average number differences within population



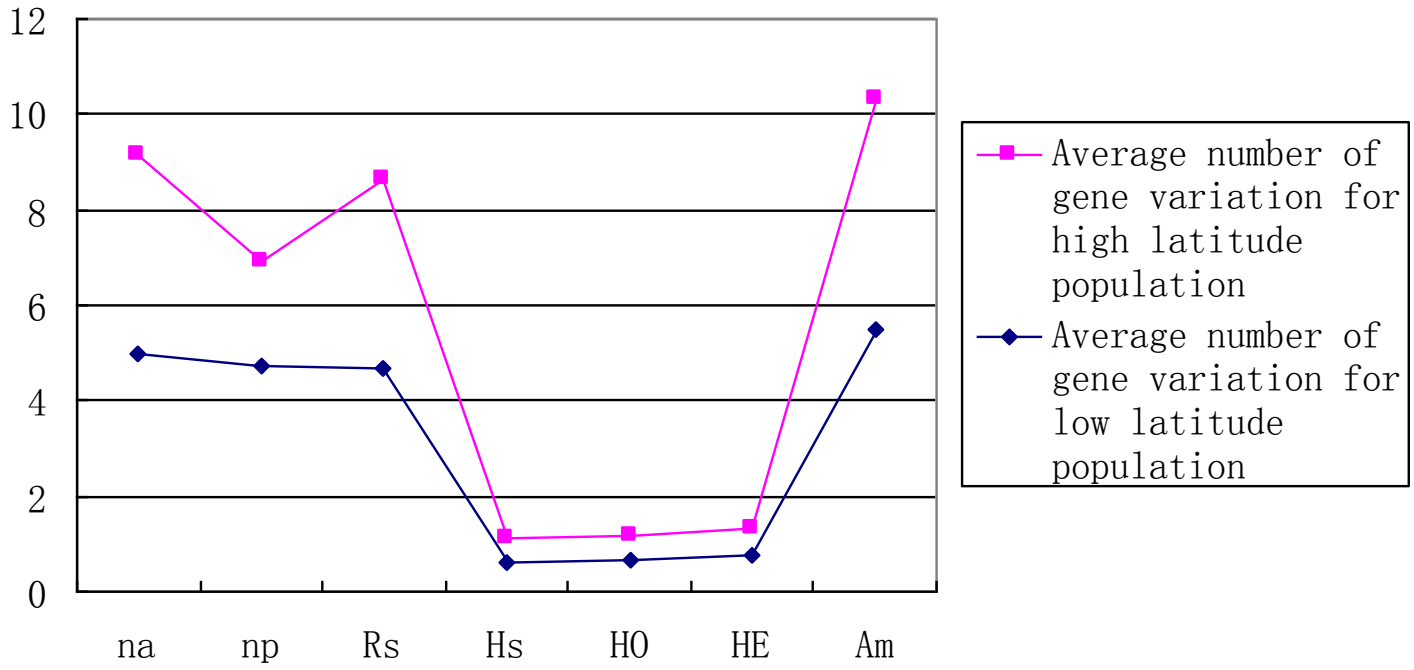


Fig.4 Comparision for the mean value of genetic variation between population of lower latitude and high latitude. Am is Average number differences within population

- ***Population structure***

  Differentiation among populations, as calculated by the fixation index *FST* (Table 1), represents the first approach to detect population structure. Pairwise *FST* values ranged from 0.046 to 0.314. Most values of *Fst* are significant excluding eleven no-significant values.

  The second way to estimate the structure of 16 populations was using the STRUCTURE2.2. In our study the ancestry of each individual is referred to one of the likely hypothetical original populations defined as clusters (K).The second derivative of the likelihood function with respect to K (ΔK), identified a break in the slope of the likelihood distribution around K=7. Therefore, 7 is the smallest value of K that captures the major structure in the data set. Table 2and Fig.5 reports the average values of ancestry probabilities (Q) of each population in the 7 clusters. In cluster1, high proportion samples from Bagan of Myanmar (Q=0.569), Siem Reap of Cambodia (Q=0.469) and Guangxi of China (Q=0.710) were assigned in this cluster. Samples of two Vietnam populations, Hanoi (Q=0.128) and Panchit (Q=0.109) also share a small part of ancestry with three populations above in cluster1. The sample from Mandala of Myanmar and Chongqing of China were primarily assigned to cluster 2. In this cluster, Panchit

- population of Vietnam shares a small part of ancestry (Q=0.166) with these two populations. Hekou and Ruili of Yunnan flies have highest probabilities of co-ancestry in cluster3. Cambodia and Myanmar-B shares a certain degree of co-ancestry with Hekou and Ruili in cluster3, with Q=0.104 and Q=0.373 respectively. In cluster 4, the highest value of ancestry is that of Guizhou, Q=0.821. The same cluster is also represented by Yunnan-Y, with Q=0.598. More than 50% membership of Myanmar-M, Vietnam-N and Vietnam-P in cluster 5 and a small probability of ancestry of Fujian and Yunnan-H are also in this cluster. Samples from Laos and Yunnan-D were primarily assigned to cluster6 and Cambodia, Yunnan-K and Yunnan-Y share a certain degree ancestry with them in cluster 6. In cluster 7, high co-ancestry values are present in Fujian population, Yunnan-Q and Yunnan-K population. In the same cluster, samples of Cambodia share a certain degree of ancestry with Q=0.112.

Table 1 Estimating Fst of Weir (1996) for each pair of populations using the ENA correction described in Chapuis and Estoup (2007)

	Myanmar-B	Myanmar-M	Vietnam –N	Vietnam -P	Cambodia	Laos	Guizhou	Fujian	Chongqing	Guangxi	Yunnan-Y	Yunnan-K	Yunnan-D	Yunnan-Q	Yunnan-H	Yunnan-R
Myanmar-B	0.000															
Myanmar-M	0.153	0.000														
Vietnam –N	0.105	0.106	0.000													
Vietnam -P	0.106	0.098ns	0.042ns	0.000												
Cambodia	0.136	0.144	0.076ns	0.077ns	0.000											
Laos	0.174	0.220	0.117	0.151	0.079ns	0.000										
Guizhou	0.189	0.212	0.135	0.147	0.100	0.088ns	0.000									
Fujian	0.135	0.186	0.124	0.124	0.085	0.114	0.124	0.000								
Wulong	0.201	0.203	0.165	0.129	0.119	0.241	0.244	0.147	0.000							
Guangxi	0.209	0.260	0.232	0.091ns	0.165	0.226	0.233	0.195	0.235	0.000						
Yunnan-Y	0.167	0.201	0.119	0.103	0.147	0.165	0.205	0.163	0.244	0.307	0.000					
Yunnan-K	0.137	0.198	0.129	0.130	0.130	0.153	0.148	0.159	0.246	0.231	0.127	0.000				
Yunnan-D	0.210	0.238	0.141	0.143	0.124	0.173	0.155	0.231	0.231	0.275	0.209	0.186	0.000			
Yunnna-Q	0.185	0.213	0.139	0.130	0.139	0.173	0.144	0.187	0.244	0.314	0.148	0.170	0.146	0.000		
Yunnan-H	0.149	0.180	0.075ns	0.052ns	0.122	0.152	0.173	0.126	0.224	0.186	0.219	0.169	0.2410	0.243	0.000	
Yunnan-R	0.046ns	0.212	0.133	0.148	0.124	0.146	0.209	0.145	0.221	0.199	0.206	0.175	0.221	0.252	0.086ns	0.000

Table 2 Average coefficients of ancestry obtained from a structure run with K = 7, for the 346 individuals of *Bactrocera dorsalis* from the 16 geographical samples. The highest value of co-ancestry of each population in a cluster is in bold

Population	Clusters						
	1	2	3	4	5	6	7
Myanmar-B	**0.569**	0.009	0.373	0.007	0.018	0.016	0.009
Myanmar-M	0.018	**0.413**	0.026	0.017	**0.502**	0.015	0.009
Vietnam -N	0.128	0.063	0.020	0.009	**0.698**	0.072	0.010
Vietnam -P	0.109	0.166	0.078	0.023	**0.539**	0.050	0.034
Cambodia_X	**0.469**	0.056	0.104	0.027	0.054	0.178	0.112
Laos_L	0.011	0.011	0.033	0.009	0.051	**0.812**	0.073
Fujian	0.037	0.015	0.016	0.019	0.118	0.032	**0.762**
Guangxi	**0.710**	0.084	0.048	0.015	0.021	0.059	0.063
Chongqing	0.015	**0.924**	0.008	0.007	0.012	0.008	0.026
Guizhou	0.057	0.035	0.014	**0.821**	0.017	0.047	0.009
Yunnan-H	0.039	0.051	**0.695**	0.007	0.101	0.056	0.051
Yunnan-R	0.030	0.024	**0.788**	0.026	0.023	0.033	0.077
Yunnan-Q	0.031	0.024	0.013	0.015	0.054	0.015	**0.847**
Yunnan-K	0.031	0.015	0.063	0.008	0.024	0.100	**0.759**
Yunnan-Y	0.013	0.018	0.015	**0.598**	0.020	0.324	0.012
Yunnan-D	0.020	0.017	0.017	0.016	0.019	**0.903**	0.007

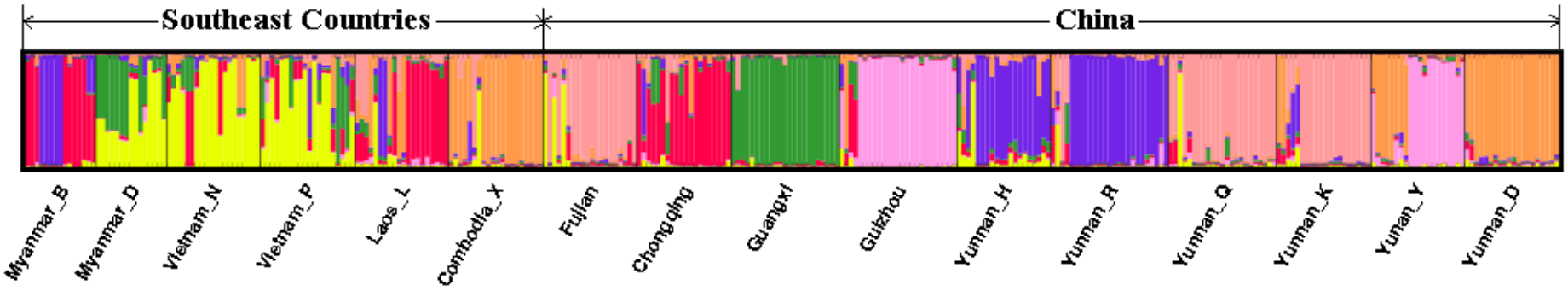


Fig. 5 Results of Bayesian clustering analysis of *B. dorsalis* individuals based on microsatellite genotypes using Structure. The seven colours (red, green, yellow, peach, orange, pink and blue) represent the co-ancestry distribution of the 346 individuals in seven hypothetical clusters (K1-K7) respectively. Bars are partitioned into seven shaded segments proportional to the inferred ancestry of each individual to each cluster

- ***Assignment rate***

Migration. Geneclass 2.0 was used to estimate the proportion of immigrants (m) into each population and the results are showed in Table 7. The diagonal values of the assignment matrix in table 8 indicate the average probability with which individuals are assigned to the corresponding reference population, the self-assignment probability values ranged from 0.348 (Mandaly) to 0.650 (Y-Kunming). Comparing all migration estimates (column values), the values ranged from 0.003 (to Bangan) to 0.425 (Hekou of Yunnan). Moreover, among 15 migration values from Hanoi to other populations, about 60% migration values are over 0.050 and about 40% migration values are over 0.100. But the opposite direction, that is from other populations to Hanoi, most of the values of migration are 0.000. The very similar tendency existed in Panchit, Cambodia and Fujian of China. Especially from Fujian to Yunnan populations, five of six migration values are above 0.050. But the opposite direction is very low. The highest value of migration is observed from two samples of Vietnam to Hekou (from Hanoi to Hekou is 0.425, from Panchit to Hekou is 0.213) and from Cambodia to Hekou (0.302).

Calculation the mean value of migration from lower latitude population to higher latitude populations, the mean value is 0.0377. Conversely, a significant lower value (0.0104) ($P<0.05$) is observed from high latitude population to low latitude populations after U test.

- ***Analysis of Bottleneck.***

  Under the two considered mutation models, SMM and TPM, recent bottlenecks and population expansion (heterozygosity deficiency) for 16 *B. dorsalis* populations has been listed in Table 9. A significant bottlenecks were detected under SMM model in Fujian population (P=0.02734) and under TPM model in Yunnan-H (P=0.03906<0.05), Yunnan-R (P=0.03594<0.05), Yunnan-D (P=0.01172<0.05) Guangxi (P=0.042578<0.05) and Guizhou populations (P=0.04469<0.05). Under SMM model, a significant probability of occurrence of a population expansion was detected in Myanmar-B, Vietnam-N, Vietnam-P, Laos-L populations. However, under the more relaxed hypothesis of an increasing proportion of mutations encompassing more than a single repeat unit (TPM), the statistical support for the population expansion signal is confirmed only for Vietnam-P (P=0.00977<0.01).

Discussion

The impact of latitude variation for gene diversity

From our data, populations from areas of latitude lower than 24° N poses a higher mean gene diversity than that of populations from areas of latitude higher than 24° N. Indeed, samples of low latitude including all southeast Asia samples and some samples of southwest and south part of China fall into the region between 24° N and 25° S, where has been describe as typical tropical climate belt with characterize of high temperature the whole year and humid (http://baike.baidu.com/view). A great number of tropical fruits, specially the mango which the favorite host of *B. dorsalis* is wide planting in these regions including Vietnam, Thailand, Guangxi, Fujian and south part of Yunnan, total more than 90000 arc mango are planted (Liang et al. 2003; http://www.cnvnlo.com). Suitable climatic condition and abundant host plants make samples of low latitude can pertain a stable and large populations, this has been demonstrated by the presence of higher number of private alleles with high average frequency, such as: Guangxi (2 private alleles with 0.088 frequency), Myanmar-B (14 private alleles with frequency of 0.098), Vietnam-N (13 private alleles with frequency of 0.044), Yunnan-Y (5 alleles with frequency of 0.109) and Fujian (3 private alleles with frequency of 0.075). Homogenous environments and stable population in low latitude areas encourage a continuous gene flow and maintain high degree gene diversity. Conversely, samples of high latitude areas (latitude higher than 24ºN) fall into regions between 24ºN and 30ºN, where most samples are belong to monsoonal climate (Wang 2002). Correlative ecological niche models (CLMEX model) and other ecological researches show some populations in this region are seasonal distribution (Hou et al. 2005, Ye 2001) such as: Yunnan-K, Yunnan-L, Yunnan-D, Yunnan-Q, Chongqing, Guizhou populations and including north part of India (Sudhakar Rao and Sundaresan, 2007), but absenting samples in this study. Theses populations always suffered from seasonal bottleneck in winter and it means unstable populations existed in these regions and also subject to reducing gene diversity.

- **New evidence for native range of *B. dorsalis***

The *B. dorsalis* was first recorded in Koshun, Taiwan (Hardy 1973; Drew and Hancock 1994), but it does not mean that Taiwan pertain the native range of *B. dorsalis*. Aketarawang compared the genetic structure of *B. dorsalis* from Far East Asia, South Asia, Southeast Asia and the Pacific Area and hypothesized either Far East Asia (China or Taiwan) or Southeast Asia could be the source areas of this pest. Finally, China has been regarded as the most probably source area of *B. dorsalis* (Aketarawong et al. 2007). But this evidence is needed to be confirmed because too less China individuals in his study.

When consider the native range of a species based on molecular data, it is important to evaluate its genetic diversity. Generally, native population will have the greatest gene diversity (Templeton 1998,; Roderick 2004). The genetic data in this study suggested samples of lower latitude in Asia with stable and large population pertain higher mean gene diversity. Some populations of low latitude had showed characterizes of native range such as: Myanmar-B, Vietnam-N and Vietnam-P with obvious higher values of microsatllite diversity. Collectively, areas of latitude lower than 24ºN in Asia most probably pertain the native range of *B. dorsalis*. The historical records show that early records of *B. dorsalis* all appeared in low latitude areas of Asia but not in higher latitude regions. For example, the first record of this pest is in Taiwan in 1912 (Li et al. 2007). Shortly afterwards (in 1916), India and Myanmar also reported the appearance of this fly (Bezzi 1916). In China, the first record of this fly was in 1934 in Hainan (Xie 1937, Wang 1996). These historical records provide evidences that *B. dorsalis* stayed a relatively long periods in low latitude regions of Asia and supported the hypothesis about the native range of the fly above.

- Two samples of Vietnam-N and Vietnam-P, separated from XX kilometers each other () and one samples of Myanmar, Bangan, are characterized by a significant numbers of alleles coupled with a high numbers of private alleles occurring at high frequency in all samples of this study. These are obvious features as native range of this species. Exactly, these samples are within the native range of the fruit fly. In a similar way, the Guangdong fly population has been described by Aketarawong with obvious features of being native range (Aketarawong et al. 2007), perhaps it also within the source areas.

Colonize pathway for *B. dorsalis* in Asia

- Under the hypothesis that areas of latitude lower than 24ºN in Asia maybe pertain the native range of *B. dorsalis*, the whole colonize pathway of the fly is needed to learn. The two methods based on individual assignment, provided by Geneclass and Structure, allowed us to infer the main pathway of *B. dorsalis.* Two main results are indicative: (i) Low latitude populations has an asymmetric mean gene flow towards high latitude populations. This asymmetric tendency also can found in specific populations, for instance, from Vietnam to Guizhou and Yunnan-Q, from Fujian to Yunnan-K, from Cambodia to Yunnan-Q. (ii) the higher rate of co-ancestry were existed between low latitude populations and high latitude populations in 7 hypothesis ancestry clusters (Table 2, Fig. 5). They are Myanmar and Chongqing, Yunnan-Y and Guizhou, Laos and Yunnan-D. In addition, many low latitude populations showed they experienced a significant population expansion including Vietnam, Fujian, Laos and Myanmar populations based on Bottleneck test. These evidences lead to the tendency of colonize pathway for this fly in Asia from south to north (from low land to high land).

The Genetic structure and immigrant routes of *B. dorsalis* in China

It is clear from structure analysis (Fig.5) that the population subdivision in China is not on a geographical basis as the south areas of China flies are not genetically separate from southwest flies. In fact, the whole China populations formed a homogenous unite and an extensive gene flow were existed among China populations.

Given the idea that a colonize pathway from low latitude areas to high latitude areas for this fly, China must be experiences being introduced from other non-China regions. Li deduced that there are two routes that the fly is introduced from Southeast Asia to China. One is from southwest part of China, another one from south part of China (Li et al. 2007). In our studying, Yunnan-R, southwest populations of China and Yunnan-H, Guangxi, two China south populations show clear features of entrance that the fly was introduced from Southeast Asia. As hypothesis one of native range, Vietnam, showed a great gene flow to Yunnan-H and Guangxi than vice versa. Moreover, a very close genetic relationship (no-significant Fst values,Table 1) was also detected between Vietnam and the two China populations. Another assumed native range of Myanmar-B show a high degree co-ancestry and high gene similarity with Yunnan-R. Furthermore, Yunnan-R, Yunnan-H and Guangxi populations all experienced obvious population bottleneck as introduced entrance. Our genetic data are in accordance with Li's assumption to some extend.