

# ajava

Asian Journal of Animal and Veterinary Advances



Academic  
Journals Inc.

[www.academicjournals.com](http://www.academicjournals.com)

## **A Method to Rapidly Identify to What Species Unknown Animals are Closely Related**

Hongwei Li, Chunjiang Zhao, Haiyue Xu, Bo Yu and Changxin Wu

College of Animal Science and Technology, China Agricultural University, Beijing, 100193, China

*Corresponding Author: Changxin Wu, College of Animal Science and Technology, China Agricultural University, Beijing, 100193, China Tel: 86-10-62733683 Fax: +86-10-62732738*

### **ABSTRACT**

Classic taxonomy of animal mainly depends on genetic, morphology, fossil and distribution studies, so the study aims at developing a method which can rapidly identify to what species unknown animals are closely related by comparison of genomic sequences. With more complete sequenced animal genomes available, genomic information will play a key role on classifying unknown animals. Here, we put forward a method how to rapidly identify unknown animals. The method includes: (1) selecting a vector which is convenient for cloning and sequencing; (2) digesting the total genomic DNA from unknown animals and the vector with the same two enzymes which have recognition sites on the vector; (3) ligating the digested DNA to the vector; (4) transforming the ligation products to competent cell; (5) selecting positive clones to sequence and blast with sequenced animal genomes; (6) then finding out the most close relative species to the unknown animals. In the paper we took a kind of unknown wild flies which captured in nature, for example, to display the method. By analysis of data from blast output, results showed that eleven query sequences (1, 2, 4, 5, 6, 7, 9, 10, 12, 13, 14) from unknown flies have the highest similarities with subject sequences on the *D. ananassae* genome, but only minor parts of No. 2, 4~7, 9, 10 and 12 sequences have similarities with subject sequences on eleven drosophila genomes and the unknown flies are the most closely related to *D. ananassae* in evolutionary contexts. it will be a promising power in rapidly identifying to what species unknown animals is closely related when more complete sequenced animal genomes are available .

**Key words:** Taxonomy, vector, ligation, the expect value, alignment score, blast

### **INTRODUCTION**

The species problem is the oldest in biology. To taxonomists, species are categories of classification. Classic taxonomy of animal mainly depends on genetic, morphology, fossil and distribution studies. Presently much biological research depends upon species diagnoses. The Consortium for the Barcode of Life (CBOL) is an international initiative devoted to developing DNA barcoding as a global standard for the identification of biological species. The Folmer region at the 5' end of the cytochrome c oxidase subunit 1 mitochondrial region (COI) is serving as the standard barcode region for almost all groups of higher animals (Folmer *et al.*, 1994; Brown *et al.*, 1999; Bucklin *et al.*, 1999; Hebert *et al.*, 2003a, b). But it is impossible for any DNA barcode to fully resolve the complexity of life. For example, where species boundaries have been blurred by hybridization or introgression, supplemental analyses of one or more nuclear genes will be required. Moreover, it will be more difficult to diagnose young species in groups with slow rates of evolution

than those with rapid rates of molecular evolution. So under these circumstances, comparable genomics (e.g., genome size and sequence similarities) will play a key role in taxonomy of animal. The BLAST programs are widely used tools for DNA databases for sequence similarities, but BLAST hits are depends on the availability of close relatives present in the databases (Koski and Golding, 2001). Twelve sequenced drosophila genomes have been available on [http:// www. flybase. org](http://www.flybase.org), so in the paper taking unknown flies, for example, to display the method how to classify unknown animals found in nature by genomic information.

The genus *Drosophila*, belonging to the family Drosophilidae and being very diverse in appearance, behavior and breeding habitat, includes over 2000 described and more undescribed species in the world (O'Grady and Markow, 2009). Typically taxonomy of species group of *Drosophila* is mainly based on genetic, morphology and distribution studies (O'Grady and Markow, 2009; Coyne, 1993). Nevertheless, Sibling species are distinguished with difficulty in morphology meanwhile different species can live in the same territory. Genetic crosses can check whether there are offspring to produce in a cross of two species, but producing no offspring could be involved in several mechanisms (Coyne, 1993). Moreover, sibling species have contrasting difference in polymorphism of chromosome (Rohde *et al.*, 2006). At the same time, there exists widely apparently difference in morphology but no alterations in chromosome among different species (Sucena and Stern, 2000). So, a growing number of studies from polymorphism of enzymes to variability of DNA sequence are making *Drosophila* taxonomists turn to detect difference at genomic level among different species (Gao *et al.*, 2007; Bosco *et al.*, 2007; McBride and Arguello, 2007; Noor *et al.*, 2007). In fact Sibling species have similarity in morphology, but very difference in genetic materials. With completion of sequenced genomes of 12 species of *Drosophila* and their rapidly extending use in comparative biology (Adams *et al.*, 2000), *Drosophila* 12 Genomes Consortium (Stark *et al.*, 2007), comparison of sequence at genome level can provide a promising power to classify the most related species to unknown flies.

## MATERIALS AND METHODS

The project was supported by the National program on key Basic Research projects of China (Grant No. 2006CB102101). It started on April, 2009, ending on December, 2009.

**Procedure of the method:** In our laboratory taking unknown flies for example, we forwarded a method how to identify unknown animals found in nature. The method includes: (1) selecting a vector suitable for cloning and sequencing; (2) selecting two enzymes which have recognition sites on the vector and share common reaction buffer to digest the total genomic DNA from unknown flies and the vector; (3) ligating the digested genome to the vector; (4) transforming the ligation to competent cell; (5) selecting positive clones to sequence and blast with twelve sequenced drosophila genomes (Fig. 1).

**DNA extraction:** An unknown wild flies were captured in Sanya, Hainan, People's Republic of China. Extraction of genomic DNA from the wild flies was followed by a protocol (Sullivan *et al.*, 2000) and then purification by phenol-chloroform extraction, precipitation by absolute ethanol, resuspended in deionized distilled water, stored at -20°C for latter digestion.

**Digestion of genomic DNA and vector:** A plasmid, PMD18T vector was selected from Takara company (Japan), which is usually used for cloning and sequencing (Fig. 2). Based on recognition

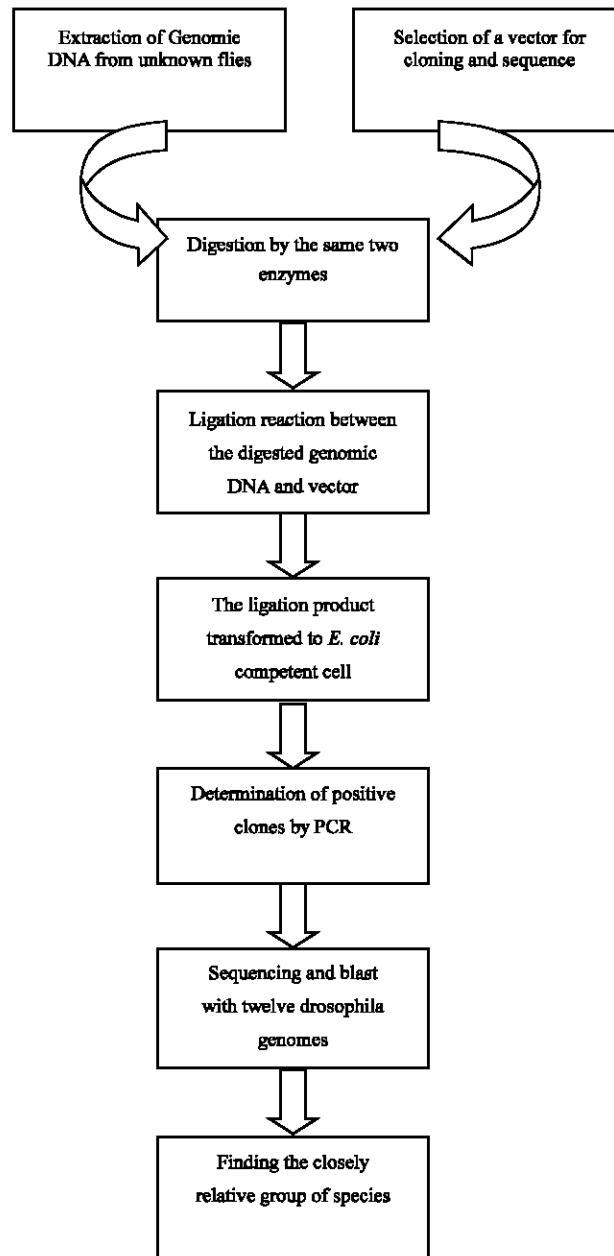


Fig. 1: Scheme of the method

sites on the vector, Pst 1 and HindIII were selected (Takara company, Japan) which have recognition sites on the vector and common reaction buffer to digest the genomic DNA and the vector. Reaction system includes: 1  $\mu$ L Pst1, 1  $\mu$ L HindIII, 2  $\mu$ L 10x Buffer, DNA $\leq$ 1  $\mu$ g, adding deionized distilled water to 20  $\mu$ L, incubated at 37°C for 3 h.

**Ligation reaction:** The digested genomic DNA and vector were purified by TIAN Quick Mini Purification Kit (TIANGEN, Beijing, China) and then the digested genomic DNA was ligated to the digested vector. A Ligation procedure was followed by the protocol from a kit (Takara, Japan) and

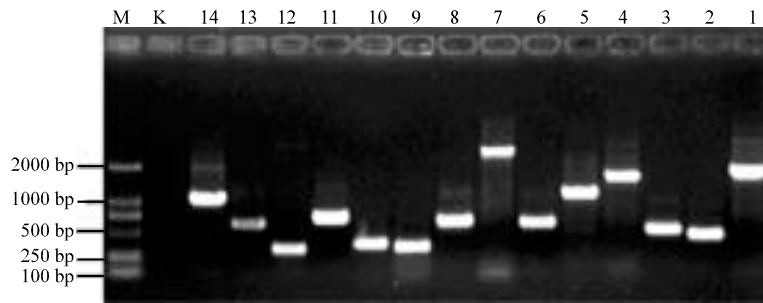


Fig. 2: PCR for identifying the positive clones. M-DNA marker, K-negative control, number :1,2,3,4,5,6,7,9,10,12,13,14 representing the selected clones for sequencing. Primers for amplification of PCR: RV-M 5' GAGCGGATAACAATTTACACAGG 3', M13-47 5' CGCCAGGGTTTTCCCAGTCACGAC 3' (Fig. 2). The PCR reaction was implemented under the following conditions:denaturing at 94°C for 4 min, 30 cycles of denaturing at 94°C for 45 sec, annealing at 60°C for 45 sec and extension at 72°C for 2 min, followed by at 72°C for 7 min. The PCR products were electrophoresed on 1.2% agarose gel and stained with EBr

then 10  $\mu$ L of the ligation reaction was transformed to 100  $\mu$ L DH5 $\alpha$  competent cell (TIANGEN, Beijing, China), followed by 30 min on ice, heated at 42°C for 90 sec, immediately put on ice for 2 min, adding 800  $\mu$ L LB liquid, shaking at 37°C for 60 min and put the liquid on solid LB medium with concentration of 100  $\mu$ g mL<sup>-1</sup> Ampicillin, incubation at 37°C for 8~12 h.

## RESULTS

Positive clones were selected to sequence and blast with twelve completed drosophila genomes. The positive clones were determined by PCR with universal primers on PMD18T vector. Twelve positive clones with different sizes (No. 1~7, 9, 10, 12, 13, 14). No. 6, 8 and 11clones have the same length, so only No. 6 was selected to sequence (Fig. 2). Sequencing was completed on ABI3703XL sequencer at Chinese National Human Genome, Beijing, China. BLOSUM62 scoring matrix was selected when blasting all sequences by BLAST (Altschul *et al.*, 1997) on www.Flybase.org/blast. The scoring matrix, generated by Henikoff and Henikoff (1992), is the most commonly used matrix for scoring protein and nucleotide sequence alignments (Mount, 2004).

**Total scores showed the unknown flies are the most closely related to *D. ananassae* in evolutionary contexts:** Similarity between sequences implies evolutionary kinship (Reich *et al.*,1984; Karlint and Altschul, 1990). Search for sequence similarity depends on two important parameters: the score of the alignment and the expect value (E) of alignment. Alignment score is computed score based on the number of matches, substitutions. Higher score indicates that there exists higher similarity between query sequence and subject sequence. Expect value (E) is the number of unrelated sequence in a similarity search of a sequence database that are expected to achieve a local alignment score as high or higher than the one obtained between the query sequence and the matching database sequence. Contrary to the alignment score, the lower number of E value is, the higher match between query sequence and subject sequence is. E values

Table 1: A hit summary of scores from blast output

Query sequence	Size	Source of subject sequence and score											
		Dmel S	Dsim S	Dsec S	Dyak S	Dere S	Dana S	Dpse S	Dper S	Dwil S	Dmoj S	Dvir S	Dgri S
1	1638						3105						
2	347	56	56	56	48	52	688		46			40	40
3	400												
4	1446	119	135	135	133	112	2558	115	54	100	72	86	44
5	2619	238	230	238	246	246	2853	149	149	129	115	123	123
6	511	121	121	121	127	135	500	96	88	50	76	54	70
7	212	68	68	76	70	70	413	84	84	40		40	42
9	249		40	40	40		254	40	40		42		40
10	384	42	40	40	52	66	544	44				42	
12	192	86	86	86	86	86	326	72	72	42	48	48	48
13	470						648						
14	907						1701						
Total	9339	738	776	792	802	767	13572	600	533	361	353	433	407

BLOSUM62 scoring matrix was selected when aligning all sequences on www.Flybase.org/blast. advanced setting: The default value (1) means that 1 such matches are expected to be found merely by chance, according to the stochastic model of Karlin and Altschul (1990). Lower EXPECT thresholds are more stringent, leading to fewer chance matches being reported. Restricts the number of short descriptions of matching sequences reported to the number specified; default limit is 25 descriptions. Restricts database sequences to the number specified for which high-scoring segment pairs (HSPs) are reported; the default limit is 25

of 0 indicate that the match is to the query sequence itself. So by analysis of BLASTN output, the results indicated that eleven query sequences (1, 2, 4, 5, 6, 7, 9, 10, 12, 13, 14) from unknown flies have the highest similarities with subject sequences on the *D. ananassae* genome, but only minor parts of No. 2, 4~7, 9, 10 and 12 sequences have similarities with subject sequences on eleven drosophila genomes (Table 1). In further analysis, three sequences including No. 1, 13 and 14 clones which have E values of 0 showed that they have complete identities with sequences from Scaffold13258, Scaffold 13034, Scaffold13230 of Dana genome, respectively (Table 3). No similarities of the three sequences, however, were found in other eleven sequenced drosophila genomes (Table 1). These Scaffolds are located in intragenic region of Dana genome, showing that the regions are highly variable in evolutionary process. Moreover, No. 2, 4~7, 9 and 10 clones, all these sequence have highly identities with subject sequences from seven genes of Dana (Table 3). Among these sequences, No. 2, 4, 5 sequences with E values of 0 showed that the three sequences have complete homologies with sequences from genes GF15054, GF23996 and GF17520 of Dana, respectively (Table 2). No. 6, 7, 9, 10 sequences having E values of  $3.2e^{-122}$ ,  $2.2e^{-113}$ ,  $1.5e^{-65}$ ,  $1.7e^{-152}$ , respectively, indicated significant similarity with subject sequences from genes GF11079, GF18158, GF16092 and GF12405 of Dana, respectively (Table 2). These results further confirmed that the unknown flies are the most closely related to *D. ananassae* in evolutionary contexts. In fact, *D. ananassae* is a cosmopolitan and circumtropical species. A distribution study also confirmed that *D. ananassae* is widely distributed in Hainan, People's China in which the unknown flies were captured (Qian *et al.*, 2006).

Table 2: A hit summary of E value from blast output

Query sequence	Size	Source of subject sequence and E value											
		Dmel	Dsim	Dsec	Dyak	Dere	Dana	Dpse	Dper	Dwil	Dmoj	Dvir	Dgri
1	1638						0						
2	347	9.8e <sup>-6</sup>	9.8e <sup>-6</sup>	9.8e <sup>-6</sup>	0.002	10 <sup>-4</sup>	0		0.009			0.58	0.58
3	400												
4	1446	3e <sup>-24</sup>	6e <sup>-29</sup>	6e <sup>-29</sup>	2e <sup>-28</sup>	8e <sup>-22</sup>	0	5e <sup>-23</sup>	10 <sup>-4</sup>	3e <sup>-18</sup>	7e <sup>-10</sup>	4e <sup>-14</sup>	3e <sup>-12</sup>
5	2619	10e <sup>-60</sup>	2e <sup>-57</sup>	10e <sup>-60</sup>	4e <sup>-62</sup>	4e <sup>-62</sup>	0	7e <sup>-33</sup>	310 <sup>-4</sup>	6e <sup>-27</sup>	9e <sup>-23</sup>	4e <sup>-25</sup>	3e <sup>-12</sup>
6	511	3e <sup>-25</sup>	3e <sup>-25</sup>	3e <sup>-25</sup>	5e <sup>-27</sup>	2e <sup>-29</sup>	3e <sup>-129</sup>	2e <sup>-17</sup>	4e <sup>-15</sup>	910 <sup>-4</sup>	2e <sup>-11</sup>	6e <sup>-5</sup>	10e <sup>-10</sup>
7	212	2e <sup>-9</sup>	2e <sup>-9</sup>	6e <sup>-12</sup>	4e <sup>-10</sup>	4e <sup>-10</sup>	2e <sup>-113</sup>	3e <sup>-14</sup>	3e <sup>-14</sup>	0.34		0.34	0.098
9	249		0.4	0.4	0.4		1.4e <sup>-65</sup>	7e <sup>-6</sup>	3e <sup>-5</sup>		0.1		0.4
10	384	0.16	0.65	0.65	210 <sup>-4</sup>	e <sup>-8</sup>	2e <sup>-152</sup>	0.04				0.16	
12	192	5.8e <sup>-15</sup>	5.8e <sup>-15</sup>	5.8e <sup>-15</sup>	5.8e <sup>-15</sup>	5.8e <sup>-15</sup>	4e <sup>-87</sup>	8.7e <sup>-11</sup>	8.7e <sup>-11</sup>	0.08	10 <sup>-3</sup>	10 <sup>-3</sup>	10 <sup>-3</sup>
13	470						0						
14	907						0						

BLOSUM62 scoring matrix was selected when aligning all sequences on www.Flybase.org/blast. advanced setting: The default value (1) means that 1 such matches are expected to be found merely by chance, according to the stochastic model of Karlin and Altschul (1990). Lower EXPECT thresholds are more stringent, leading to fewer chance matches being reported. Restricts the number of short descriptions of matching sequences reported to the number specified; default limit is 25 descriptions. Restricts database sequences to the number specified for which high-scoring segment pairs (HSPs) are reported; the default limit is 25

Table 3: A summary of identities among sequences from unknown flies with Dana genome

Query sequence	Length	Subject sequence	Species	Identities
1	1638	Scaffold13258	<i>D. ana.</i>	1617/1638(98.7%)
2	347	GF15054	<i>D. ana.</i>	347/347(100%)
3	400	Genome	<i>Wolbachia</i>	72%
4	1446	GF23996	<i>D. ana.</i>	1420/1446(98.2%)
5	2619	GF17520	<i>D. ana.</i>	2533/2619(96.7%)
6	511	GF11079	<i>D. ana.</i>	456/511(89.2%)
7	212	GF18158	<i>D. ana.</i>	211/212(99.5%)
9	249	GF16092	<i>D. ana.</i>	216/249(86.7%)
10	384	GF12405	<i>D. ana.</i>	301/384(78.3%)
12	192	Scaffold13266	<i>D. ana.</i>	183/192(95.3%)
13	470	Scaffold13034	<i>D. ana.</i>	422/470(89.8%)
14	907	Scaffold13230	<i>D. ana.</i>	893/907(98.5%)

BLOSUM62 scoring matrix was selected when aligning all sequences on www.Flybase.org/blast. advanced setting : The default value (1) means that 1 such matches are expected to be found merely by chance, according to the stochastic model of Karlin and Altschul (1990). Lower EXPECT thresholds are more stringent, leading to fewer chance matches being reported. Restricts the number of short descriptions of matching sequences reported to the number specified; default limit is 25 descriptions. Restricts database sequences to the number specified for which high-scoring segment pairs (HSPs) are reported; the default limit is 25

## DISCUSSION

Interestingly similarity of No. 3 sequence is not found in twelve drosophila genomes (Table 1), but has highly identical with sequence from bacteria, *Wolbachia* genome, which often lives on drosophila (Table 3). Endosymbionts, such as *Wolbachia* to host transfers were found in three sequenced insect genomes (Hotopp *et al.*, 2007). This implied that No. 3 sequence transfer from the *Wolbachia* genome to the unknown drosophila chromosome. In previous paper instances of gene transfer have been reported by Lawrence and Ochman (1998), Nelson *et al.* (1999) and Eisen

(2000). So, the caution should be taken when using sequence similarity to infer evolutionary relationships (Sicheritz-Ponten and Anderson, 2001; Eisen, 1998; Eisen and Hanawalt, 1999).

The BLAST programs are widely used tools for searching protein and DNA databases for sequence similarities. Similarity between or within sequences suggests evolutionary relationship (Reich *et al.*, 1984; Karlint and Altschul, 1990; Altschul *et al.*, 1997; Mount, 2004). High sequences similarity between two species implies a closely relationship in evolutionary context. So, by analysis of BLAST hits, one can find out to what species unknown animals is closely related when more complete sequenced animal genomes are available. The method is easy and effective, but its effectiveness depends on BLAST hit. The more complete the data base the better BLAST will work. Interestingly, non-coding genomic sequences will play more important roles on identifying unknown animals found in nature than those of coding sequences in that their character of highly variation can directly point to the most closely related species (Table 1).

Presently there exist many unknown animals in the world, which requires taxonomists to classify them. By classic taxonomy it is a difficult task to rapidly identify unknown animal species when there exists hybridization, young species and slow molecular evolution, but using genomic information will overcome the disadvantages to complete classification of unknown animals.

## CONCLUSION

The above processes display a powerful method how to identify unknown animals. The method only needs digestion of genomic DNA by restriction nucleases, ligation of vector, sequence analysis and blast with sequenced animal genomes. More importantly, most labs meet these requirements to manipulate the method, so it will be a very helpful tool for taxonomists to classify unknown animals when more complete sequenced animal genomes are available.

## REFERENCES

- Adams, M.D., S.E. Celniker, R.A. Holt, C.A. Evans and J.D. Gocayne *et al.*, 2000. The genome sequence of *Drosophila melanogaster*. *Science*, 287: 2185-2195.
- Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, W. Miller and D.J. Lipman, 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl. Acids Res.*, 25: 3389-3402.
- Bosco, G., P. Campbell, J.T. Leiva-Neto and T.A. Markow, 2007. Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. *Genetics*, 177: 1277-1290.
- Brown, B., R.M. Emberson and A.M. Paterson, 1999. Mitochondrial COI and II provide useful markers for *Weiseana* (Lepidoptera, Hepialidae) species identification. *Bull. Entomol. Res.*, 89: 287-294.
- Bucklin, A., M. Guarnieri, R.S. Hill, A.M. Bentley and S. Kaartvedt, 1999. Taxonomic and systematic assessment of planktonic copepods using mitochondrial COI sequence variation and competitive, species-specific PCR. *Hydrobiology*, 401: 239-254.
- Coyne, J.A., 1993. The genetics of an isolating mechanism between two sibling species of *drosophila*. *Evolution*, 47: 778-788.
- Eisen, J.A., 1998. Phylogenomics: Improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.*, 8: 163-167.
- Eisen, J.A. and P.C. Hanawalt, 1999. A phylogenomic study of DNA repair genes, proteins and processes. *Mutat. Res.*, 435: 171-213.



- Eisen, J.A., 2000. Horizontal gene transfer among microbial genomes: New insights from complete genome analysis. *Curr. Opin. Genet. Dev.*, 10: 606-611.
- Folmer, O., M. Black, W. Hoeh, R. Lutz and R. Vrijenhoek, 1994. DNA primers for amplification of mitochondrial cytochrome oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.*, 3: 294-299.
- Gao, J.J., H.A. Watabe, T. Aotsuka, J.F. Pang and Y.P. Zhang, 2007. Molecular phylogeny of the *Drosophila obscura* species group, with emphasis on the old world species. *BMC Evolut. Biol.*, 7: 87-87.
- Hebert, P.D.N., A. Cywinska, S.L. Ball and J.R. de Waard, 2003a. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B*, 270: 313-321.
- Hebert, P.D.N., S. Ratnasingham and J.R. de Waard, 2003b. Barcoding animal life: Cytochrome oxidase subunit 1 divergences among closely related species. *Proc. Biol. Sci.*, 270: S596-S599.
- Henikoff, S. and J.G. Henikoff, 1992. Amino acid substitution matrices from protein blocks. *Proc. Nat. Acad. Sci. USA.*, 89: 10915-10919.
- Hotopp, J.C., M.E. Clark, D.C.S.G. Oliveira, J.M. Foster and P. Fischer *et al.*, 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science*, 317: 1753-1756.
- Karlint, S. and S.F. Altschul, 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. USA.*, 87: 2264-2268.
- Koski, L.B. and G.B. Golding, 2001. The closest BLAST hit is often not the nearest neighbor. *J. Mol. Evol.*, 52: 540-542.
- Lawrence, J.G. and H. Ochman, 1998. Molecular archaeology of the *Escherichia coli* genome. *Proc. Nat. Acad. Sci. USA.*, 95: 9413-9417.
- Mcbride, C.S. and J.R. Arguello, 2007. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics*, 177: 1395-1416.
- Mount, D.W., 2004. *Bioinformatics: Sequence and Genome Analysis*. 2nd Edn., Genetics and Genome Science Biotechnology, Laboratory Manuals/Handbooks, University of Arizona, Tucson.
- Nelson, K.E., R.A. Clayton, S.R. Gill, M.L. Gwinn and R.J. Dodson *et al.*, 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature*, 399: 323-329.
- Noor, M.A.F., D.A. Garfield, S.W. Schaeffer and C.A. Machado, 2007. Divergence between the *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to chromosomal inversions. *Genetics*, 177: 1417-1428.
- O'Grady, P.M. and T.A. Markow, 2009. Phylogenetic taxonomy in *Drosophila*. *Fly*, 3: 10-14.
- Qian, Y.H., Y.L. Liu, S.T. Li, Y. Yang and Q.T. Zeng, 2006. Compact and distribution of the *Drosophila melanogaster* species group from China. *J. Hubei Univ. (Natural Science)*, 28: 397-402.
- Reich, J.G., H. Drabsch and A. Diumler, 1984. On the statistical assessment of similarities in DNA sequences. *Nucleic Acids Res.*, 12: 5529-5543.
- Rohde, C., A.C. Garcia, V.H. Valiati and V.L. Valente, 2006. Chromosomal evolution of sibling species of the *Drosophila willistoni* group. I. Chromosomal arm IIR (Muller's element B). *Genetica*, 126: 77-88.

- Sicheritz-Ponten, T. and G.E. Anderson, 2001. A phylogenomic approach to microbial evolution. *Nucl. Acids Res.*, 29: 545-552.
- Stark, A., M.F. Lin, P. Kheradpour, J.S. Pedersen and L. Parts *et al.*, 2007. Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature*, 450: 219-232.
- Sucena, E. and D.L. Stern, 2000. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by *cis*-regulatory evolution of *ovo/shaven-baby*. *Proc. Nat. Acad. Sci. USA.*, 97: 4530-4534.
- Sullivan, W., M. Ashburner and S. Hawley, 2000. *Drosophila* Protocols. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, ISBN: 0879695862.