國立臺灣大學生命科學院動物學研究所

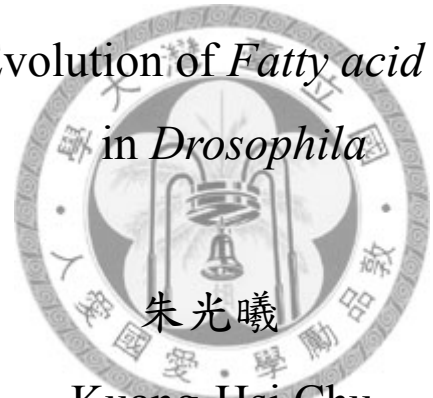
碩士論文

Institute of Zoology
College of Life Science
National Taiwan University

Master Thesis

果蠅脂肪酸去飽和酶 2 的分子演化

Molecular Evolution of *Fatty acid desaturase 2* in *Drosophila*



朱光曦

Kuang-Hsi Chu

指導教授：丁照棣
方淑

Advisors: Chau-Ti Ting
Shu Fang

中華民國 97 年 7 月

July, 2008


國立臺灣大學
教育部獎勵學生愛國愛人

Acknowledgements

我個人非常喜歡閱讀，生活中其實任何事物都可以以書來做為比擬：今天是章節多少，劇情如何發展。假如要用書來比喻我的碩士生涯的話，我想史蒂芬.金的黑塔系列 The Dark Tower (serious)(皇冠出版)應該是再合適不過。書裡描述故事的主角羅蘭如何在業(Ka)的驅動下展開尋找支撐世界的樑柱－黑塔的旅程。中文版的第一集在 2007 年 7 月出版，剛好是我碩二要瘋狂的為實驗而貢獻生命的時候；最後一集則是在 2008 年 6 月發行，而我當時正好準備要口試了。我還記得當我看到書的結尾，羅蘭在歷經千辛萬苦終於抵達黑塔時，不禁讓我潸然淚下，因為黑塔與我一直在累積的 Data 實在是太像了(念念看)。

如同許多偉大的旅程，中途一定有許多導師傾囊相助。在這兩年的碩士旅程，我必須要感謝三位老師，不只是實驗上的指導，還有許多平時生活上的建議:丁照棣老師，方淑老師以及曹順成老師。丁老師和方老師是我的指導教授，她們從我初進實驗室便給予莫大的幫助，無論是實驗上還是生活上。討論實驗的時候也會和我進行充分的腦力激盪，讓我能夠擁有科學的思維與態度(不過我還需繼續努力)。而我最感謝她們也是感到最不好意思的，就是關於英文寫作，每次都要花很多時間改我的摘要和壁報，當然這本論文就更不用說了，方老師遠在美國都還特地用數位相機把她對論文各章節的意見拍下來寄給我。而曹老師在我於中研院做實驗的過程裡給予非常多的幫助，曹老師會用相當淺顯易懂的方式很幽默的表達他的意見，請教他任何問題都可以得到滿意的答覆，還有就是關於 PCR 實驗的問

題，曹老師在這方面給予我相當大的幫助，這本論文幾乎就是用 PCR 堆起來的，所以假如沒有老師的鼎力相助，我恐怕會被這些問題搞得焦頭爛額。而與這三位老師同時進行討論更是有如醍醐灌頂，每每都感覺自己是如此的渺小，但是又何其有幸能夠站在這三位巨人的肩上。還有要感謝昆蟲系的張慧羽老師，每次在 meeting 的時候都給我很多很棒的意見。尤其是在我口試的時候，讓我能對我的研究有新的思考方向並且糾正以往偏頗的立場。還有中研院的博士後研究員:來自印度的 Paras，法國的 Stephane，日本的三代隆洋，還有王琄嬋博士。他們對我的實驗都給予許多的意見與幫助並且毫不吝嗇與我分享他們的經驗。還有我的論文口試委員：王弘毅老師，張慧羽老師，胡哲明老師和蔡玉真老師，非常感謝你們在口試時給我的建議。

在書中羅蘭擁有數位共業的夥伴(Ka-tet)，而我也很榮幸擁有好幾位實驗室的夥伴在這兩年給予我莫大的幫助。首先是小春，她扮演的是亦師亦友的角色。內建 Google 搜尋引擎與台灣美食圖鑑，我想不到有什麼事情和好吃的餐廳是她所不知道的。除了常常跟我討論實驗外，也常依據豐富的人生經驗給予我日常生活上的指引，還有要特別感謝她的是，對於我的摘要與壁報的幫忙，是的，她也是我恐怖的摘要和壁報的長期受害者之一。再來是雅仁，我跟她相處的時間並不多，所以關係並沒有很熟絡，但是她仍然很熱心的幫我看論文，並給予許多改進的建議，沒能跟她相處得更久一點，算是我碩士這兩年的最大的遺憾之一。小菊算是搞笑派的角色，他在我初進實驗室和即將要離開實驗室的兩大關鍵期扮演了很重

要的地位，他在我碩一的時候曾經把我的筆記本撕掉、果蠅 stocks 拿去急凍、弄壞我的 kit…無所不用其極打壓新人(他是無心的，我一直都這麼告訴自己)，但是他卻在忙著準備 GRE 考試的期間還抽空幫我檢查論文。還有承叡，他非常優秀，他所進行的實驗和我的很有關連，每次他的結果都讓我有新的研究方向可以思考，他即將到美國念書，希望他能順利的發展。還有跟我一樣要畢業的筱涵，非常單純善良，每次跟她聊天都會讓人非常輕鬆愉快。我的學妹莞儒大小姐和姵君以及台大的助理文喬，大馬，麒鳴，士又和其駿，他們所有人在有意無意間都幫了我不少忙，還有已經離職的小笨和佳瑋，她們常以過來人的身分教給我許多碩士生存法則。中研院果蠅研究室的各位助理也對我助益良多，尤其是祺筠，他幫我檢查整本論文，還有珈蓉，民濤，圓通，婉婷和雲露。還有已經離開中研院的家宣，嘉琳和立婷，他們教導了我許多果蠅實驗技巧。還要感謝巧雯，常常在實驗室傻呼呼的跑來跑去，逗大家開心。以及來中研院果蠅研究室當過實習生的大同大學的數位同學。張慧羽老師實驗室的小斌，嵩翰與岱樺也常常和我互相交換心得與討論各自的研究進度。他們所有人豐富了我的研究生涯，在這段漫長的時光中伴我一起成長與茁壯。

特別感謝輔仁大學生命科學系九十五年畢的某團體，其成員不時的在 MSN 上給予我鼓勵並且和我嘴砲，讓我能隨時保持向前衝的動力。

最後要感謝我的家人，他們雖然不清楚我每天在幹嘛，但是關心之情往往溢於言表；還有我的母親，她在我兩年前甫上研究所之際離我而去，但是我卻感覺

她在這段日子裡從未離開過，一直在我身邊給予支持與祝福。

在研究所的日子裡，我最大的收穫不只是完成了一項研究，還有在尋求真理的過程中所必須具備的邏輯的態度以及思維，這不光只是對於碩士有所幫助，而是對我的未來同樣有無限的影響。如同在黑塔一書的最後：羅蘭發現自己不但找到了旅行的終點，同時也找到了起點。人生充滿了一座座的黑塔，我們窮其一生不斷的重覆追尋它，但是每次我們都帶著更多的信心與勇氣踏出前進的步伐。

"I believe that when we leave a place…

…part of it goes with us, and part of it remains.

Go anywhere in the station when it is quiet and just listen.

After a while you will hear the echoes of all our conversations…

…every thought and word we've exchanged.

Long after we have gone, our voices will linger in these walls."

By G'Kar, Babylon 5: The Lost Tales.

摘要

去飽和酶 1，去飽和酶 2 和脂肪酸去飽和酶 2 在具性費洛蒙功用的果蠅表皮碳氫化合物的合成過程中扮演重要的角色。而脂肪酸去飽和酶 2 在黃果蠅和賽昔爾果蠅中參與合成雌性所特有的費洛蒙中的二烯化合物。但是我們仍然不了解它在其他種果蠅中所扮演的角色。我在十二種果蠅的基因體中找出脂肪酸去飽和酶 2 可能的同源基因並加以分析。而結果顯示脂肪酸去飽和酶 2 可能是經由反轉錄移位所產生的基因，產生的時間點可能在水果果蠅和果蠅亞屬分化之前，但在後者中失去。我也發現脂肪酸去飽和酶 2 在黃果蠅種亞群中和雌雄二型性表皮碳氫化合物有直接關連，此基因只會表現在具有二型性表皮碳氫化合物的果蠅；在單型性果蠅則沒有。為了要研究脂肪酸去飽和酶 2 在雌雄單型性表皮碳氫化合物的果蠅中是否有發生退化現象，我分析了擬黃果蠅，亞庫巴果蠅和德氏果蠅的族群，結果在二十個擬黃果蠅的品系中找到一個品系有二十五個鹼基對的缺失，在十六個德氏果蠅的品系內發現其中四個品系帶有無義突變，而在二十個亞庫巴果蠅品系中，有十四種插入和刪除的序列缺失，導致只有三個品系具有完整的編碼序列。脂肪酸去飽和酶 2 在三種果蠅中有不同程度的突變累積，代表它在各種果蠅中的退化可能是獨立發生的。脂肪酸去飽和酶 2 提供了一個研究基因重複的分子演化的模式。

關鍵字： 表皮碳氫化合物，脂肪酸去飽和酶 2，反轉錄移位，性雙型


國立臺灣大學
人愛國愛學勵品教

Abstract

Three fatty acid desaturase genes, namely *desat1*, *desat2* and *Fad2*, in *Drosophila* involve in the production of cuticular hydrocarbons (CHs) that serve as sex pheromones. Among them, *Fad2* is responsible for female-specific diene pheromone biosynthesis in *D. melanogaster* and *D. sechellia*. However, the role of *Fad2* in other *Drosophila* species is still unknown. To address this question, the *Fad2* homologs in 12 *Drosophila* species were analyzed. Based on the sequence information, *Fad2* is originated by retrotransposition which took place before the split of *Drosophila* and *Sophophora* and subsequently lost in the *Drosophila* lineage. Furthermore, in *D. melanogaster* species subgroup, *Fad2* is expressed only in the species with dimorphic CHs. These results suggest that *Fad2* is responsible for sexually dimorphic CHs in the *D. melanogaster* species subgroup. To understand whether sequence degeneration also occurred in the non-expressed *Fad2*, populations of three CH monomorphic species, *i.e.*, *D. simulans*, *D. teissieri*, and *D. yakuba,* were surveyed. A 25-bp deletion was observed in one of 20 *D. simulans* alleles. One non-sense mutation occurred in 4 of 16 *D. teissieri* sequences. In *D. yakuba*, 14 deletions and non-sense mutation resulted in that only 3 out of 20 samples keep full-length *Fad2*. The different levels of mutation accumulation indicate that the degeneration of *Fad2* happened independently in each species. In summary, *Fad2* provides a good model to study the molecular evolution after gene duplication.

Key words: cuticular hydrocarbons, *Fad2,* retrotransposition, sexually dimorphism


國立臺灣大學
人愛國愛學勵品教

Contents


國立臺灣大學
教育愛國・學術勵品

List of Figures


國立臺灣大學
人愛國愛學勵品教

List of Tables


國立臺灣大學
教育學程・愛國愛人

Chapter 1

General Introduction

Specific mate recognition system (SMRS) consists of a set of signals and responses between potential mates. Every species has its unique SMRS that controls the information exchanged between both sexes during interplay (Patterson 1985). SMRS plays a very important role in animal courtship and mating behavior, because individuals do not respond to inappropriate signals (Greenspan and Ferveur 2000).

In *D. melanogaster*, the cuticular hydrocarbons serve as contact pheromones, female-specific 7,11-heptacosadiene can stimulate wing vibrations in males (Antony and Jallon 1982; Antony et al. 1985; Ferveur and Sureau 1996). The 7,11-dienes (fatty acid with two double bonds) are sex pheromone in *D. melanogaster*, though immature males with little or no 7,11-diens can also induce courtship behavior of mature males (Jallon and Hotta 1979; Pechine, Antony, and Jallon 1988). The study of moth pheromone showed that there were three key characteristics in the biosynthesis pathway: desaturation, chain-shortening and formation of functional group. Sex-specific pheromone was produced through these steps (Jurenka 2003). The production of sex-specific CHs pheromone in *Drosophila* might through similar steps, except it uses chain-elongation instead of chain-shortening. The study of the cuticular hydrocarbons biosynthesis pathway in *D. melanogaster* (Fig.1.1) suggested that males and females

share the same elongation-decarboxylation mechanism (Bjostad, Wolf, and Roelofs 1987; Pennanec'h et al. 1991; Reed et al. 1994; Reed et al. 1995), and several Δ9 *desaturases* are involved in the pathway (Pennanec'h et al. 1997).

Fatty acid desaturases are nonheme iron-containing enzymes that introduce a double bond to specific positions of fatty-acyl carbon chains. Delta-x (Δx) means the double bond is located on the *x*th carbon–carbon bond, counting from the carboxylic acid end and omega-x (ωx) means the double bond is located on the *x*th carbon–carbon bond, counting from the terminal methyl carbon. Some Mammals use Δ5 and Δ6 *desaturases* to synthesize highly unsaturated fatty acids (Sprecher 2000), and both *desaturases* were widely expressed in human tissues (Cho, Nakamura, and Clarke 1999a; 1999b). Some plants, the synthesis of linoleic acid and α-linolenic acid from stearic acid are catalyzed by acyl-ACP Δ9, Δ12 and *ω*3 *desaturases* (Higashi and Murata 1993; Tocher, Leaver, and Hodgson 1998). In insects, the study on *Trichoplusia ni* and *S. littoralis* showed that their Δ11 *desaturases* presented in pheromone glands had conserved biochemical properties and structures with Δ9 *desaturases* in animals (Wolf and Roelofs 1986; Rodriguez et al. 1992). There are many active *desaturases* in pheromone glands, which have specific functions in pheromone production. The *desaturases* in moth's sex pheromone biosynthesis pathway have different positional and stereochemical specificities, producing fatty acids with *Z*9, *E*9, *Z*10, *Z*11, *E*11, *Z*12,

*E*12, *Z*14 and *E*14 double bonds (Jallon and Wicker-Thomas 2003). Some *desaturases* have been proved to possess multiple functions. For example, *Bombyx mori* and *Spodoptera littoralis* had Δ11 *desaturase* with Δ10, 12 utilities (Moto et al. 2004; Serra et al. 2006); *Thaumetopoea pityocampa* contained bifunctional Δ11/Δ13 *desaturase* with acetylenase activity (Serra et al. 2007).

Eight *desaturase* genes, *desaturase 1* (*desat1*), *desaturase 2* (*desat2*), *fatty acid desaturase 2* (*Fad2*), *CG8630*, *CG9743*, *CG9747*,*CG15531* and *infertile crescent* have been identified by comparing the amino acid sequences similarity with *desaturases* in rat and yeast, and they are all on the third chromosome of *D. melanogaster* except *infertile crescent,* which locates on the second chromosome. *desat1* and *desat2* use saturated fatty acids as substrates; *Fad2*, *CG8630*, *CG9743* and *CG9747* use unsaturated ones and they all are expressed in adult flies. The function of *CG15531* is still unknown. *Fad2* is involved in the female-specific CHs biosynthesis pathway in *D. melanogaster* and *D. sechellia* (Chertemps et al. 2006). The gene *infertile crescent*, located on the 2nd chromosome were found to be expressed mainly in males (reviewed in Jallon and Wicker-Thomas 2003).

desat1 has been identified based on its highly conserved protein products in carps, rats, ticks, and yeasts. Its peptide sequence was 29% and 43% identical with *desaturase* gene products in rat and yeast, respectively (Wicker-Thomas, Henriet, and Dallerac

1997). It can be expressed in both sexes with a higher expression level in females (Lee, Pariza, and Ntambi 1996). It encodes a Δ9 *desaturase* and prefers to use palmitic acid (chain length of 16 carbons) as substrate to produce ω7 unsaturated fatty acids (Dallerac et al. 2000). *desat1* contains five exons with alternative transcription start sites which produce five different transcripts (Jallon and Wicker-Thomas 2003). The interruption of *desat1* transcription by insertion of *P*-element would reduce the overall production of unsaturated cuticular hydrocarbons in both sexes, with the effect was more pronounced in males. The decrease of unsaturated hydrocarbons is accompanied by an increase of saturated ones, suggesting that *desat1* is involved in the first desaturation step (Labeur, Dallerac, and Wicker-Thomas 2002).

Another Δ9 desaturase in *D. melanogaster* is *desat2*. The chromosome location of *desat2* is close to *desat1* and shows 65% sequence identity between them. *desat2* also encodes a Δ9 desaturase and prefers to use myristic acid (chain length of 14 carbons)to produce ω5 unsaturated fatty acids (Dallerac et al. 2000). The analysis on promoter region of *desat2* in different *D. melanogaster* strains, including African and cosmopolitan strains, showed that a 16 bp deletion near the putative transcription start site resulted in the loss of expression in cosmopolitan strains (Dallerac et al. 2000; Takahashi et al. 2001; Jallon and Wicker-Thomas 2003). So *desat2* is expressed in both sexes of *D. melanogaster* African strain and other *D. melanogaster* subgroup species,

but not in *D. melanogaster* cosmopolitan strains (Greenberg et al. 2006). The main CHs in African and Caribbean population females is 5,9-HD, whereas the main CHs in most *D. melanogaster* population females is 7,11-HD (Jallon and Pechiné 1989).

There is an asymmetric reproductive isolation between African and cosmopolitan strains in *D. melanogaster*: The Z type female prefer to mate with the Z type male and has lower or no desire to mate with the cosmopolitan strain male and this sexual isolation probably is in an incipient speciation (Wu et al. 1995). They also had different copulation latency: The M type female with high amount of 7,11-HD mated more rapidly than the female with 5,9-HD (Ferveur et al. 1996). The functional and nonfunctional *desat2* determined the females' main CHs, which were thought to be sex pheromone. The analysis of d_N/d_S ratio (the numbers of synonymous and non-synonymous substitutions per site) also showed the *desat2* had higher evolution rate than *desat1*. *desat2* was therefore suggested to involve in the differentiation of Z and M type in *D. melanogaster* (Greenberg et al. 2006).

The definition of species in biological species concept is "groups of interbreeding natural populations that are reproductively isolated from other such groups" (Mayr 1963). According to this definition, the barriers of genetic exchange must play an important role in speciation. Reproductive isolation can have diverse mechanisms, including pre- and post-mating isolation that prevents gene flow among populations.

The behavioral barrier is the most important barrier among the pre-mating mechanism (Coyne and Orr 2004). Since the CHs in *Drosophila* serve as sex pheromone and strongly affect courtship behavior, the CHs difference between species is a component of reproductive isolation leading to speciation (Coyne and Charlesworth 1997; Takahashi et al. 2001).

D. melanogaster shows sexually dimorphic CHs: The male has a large amount of 7-T and female with 7,11-HD. The second double bond in females requires an additional desaturation step that is not involved in male CHs biosynthesis pathway. *Fad2* had been found to be responsible for diene production in females of *D. melanogaster* and *D. sechellia*. *Fad2* was not expressed in both sexes of *D. simulans* that has monoenes as main CHs (Chertemps et al. 2006). The interspecific difference in major female CHs is mainly controlled by a few genes on the third chromosome (Coyne 1996). A recent study showed that the difference of CHs between females in *D. melanogaster* and *D. simulans* was due to the female-specific expression of *Fad2* and *elongaseF*. The latter was a female-specific *elongase* in *D. melanogaster* and *D. sechellia* (Chertemps et al. 2006; Chertemps et al. 2007). According to these studies, the evolution of *Fad2* might be distinct in different *Drosophila* species. *Fad2* is an intronless, a characteristic of the product of retrotransporsition (Bai et al. 2007). However, the role of *Fad2* in other *Drosophila* and the origin of *Fad2* are still unknown.

When does the retrotransposition happen in the evolution of *Drosophila*? What are the expression patterns of *Fad2* in other *Drosophila* species? What is evolution of *Fad2* in *Drosophila*? Another important issue is the pseudogenization of non-expressed *Fad2* in *Drosophila* species with sexually monomorphic CHs, such as *D. simulans*.

In chapter 2, I first searched for the homologs of *desaturase* genes in *Drosophila* and tried to reconstruct the phylogeny which can also reveal the origin of *Fad2*. Then the expression patterns of *Fad2* in *Drosophila* were described. I also discussed their function in CHs synthesis. In chapter 3, I surveyed *Fad2* in *Drosophila* species with non-functional *Fad2* to investigate evolutionary patterns during pseudogenization. These results would tell us the evolutionary fates of duplicated gene in different *Drosophila* species.

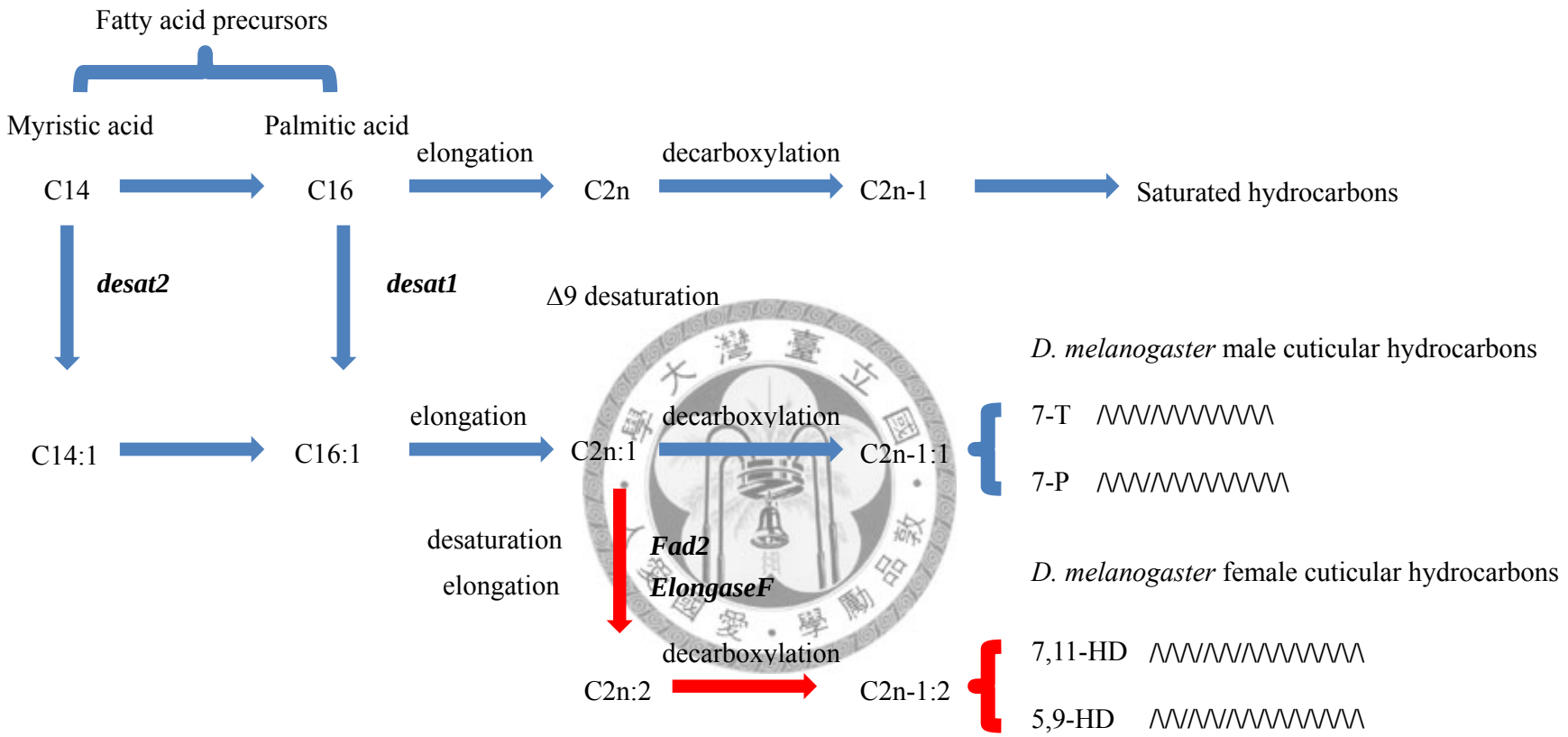


Fig 1.1. The CHs biosynthesis pathway in *D. melanogaster* (redraw from Legendre et al. 2008). The pathway used in both sexes is shown as blue lines and female–specific pathway is red.

Chapter 2

The evolution of *Fad2* in *Drosophila*

2.1.

Introduction

The *desaturase* genes play an important role in the sex pheromone production in insect. *Desaturases* are particularly significant in the generation of structural diversity of sex pheromone components, since these enzymes have evolved diverse substrate specificities, regiospecificities, and stereospecificities to produce unsaturated fatty acid precursors with a range of chain lengths, variable positions and numbers of double bonds, and both *Z* and *E* double-bond geometries (Knipple et al. 2002). As a consequence, studying the evolution of *desaturases* can help us to understand how signal divergence has occurred in the face of strong selection pressures against small changes in the signal (Roelofs and Rooney 2003). Current knowledge of the evolution of *desaturases* came from studies on moths.

The *desaturase* genes in the moth pheromone glands might have evolved from genes involved in normal fatty acid metabolism (Dugdale 1997). The *desaturase* multigene family is composed of at least four gene clusters, namely Δ9, Δ10, Δ11 and Δ14 *desaturase* genes, evolved at different rates due to different functional constrain. The Δ9 (Z9-16:Acid > Z9-18:Acid) group, which contains metabolic *desaturases* had

the slowest evolution rate, was thought to retain the ancestral gene functions. The Δ9 (Z9-16:Acid < Z9-18:Acid) and Δ10,11 groups are composed of *desaturase* genes that are involved in sex pheromone biosynthesis and evolved at faster rates because of a functional shift from metabolism to reproduction. The Δ14 group, including functional and nonfunctional *desaturase* genes in sex pheromone production possessed the fastest evolution rate among the four groups. The evolution of *desaturase* multigene family has been shown to fit the birth-and-death model (Hughes and Nei 1989; Nei and Hughes 1992; Nei, Gu, and Sitnikova 1997; Rooney, Piontkivska, and Nei 2002). Those gene families were created by gene duplication during the evolution processes, in which some genes persisted, and some were deleted or degenerated into pseudogenes. There were three characteristics which would be observed in the phylogeny of these *desaturase* gene families: first, sequences were clustered together by genes or duplication orders not by species; second, the similarity between different genes were low; third, the signatures of gene loss, deletion, and degeneration would be observed (Roelofs and Rooney 2003).

Desaturases has been proved in some *Drosophila* species to be involved in sex-specific steps in CH biosynthetic pathway (Jallon and Wicker-Thomas 2003; Chertemps et al. 2006), and played an important role in courtship behavior of *D. melanogaster* as its expression is responsible for female-specieifc CHs (Chertemps et al.

2006; Legendre et al. 2008). Genetic sequence of a region on the third chromosome, located through genetic studies and screening of deficiencies, was found to be responsible for female CHs production (Coyne 1996; Wicker-Thomas and Jallon 2000; Gleason et al. 2005). A putative *desaturase* gene *Fatty acid desaturase 2* (*Fad2*), whose peptide sequence is 53% and 54% identical with those of *desat1* and *desat2,* was detected. The putative transmembrane domains and histidine boxes were also found to be highly conserved among the three genes. Using RNA interference to knockdown *Fad2* expression in *D. melanogaster* females significantly increased the monoenes level and decrease the diene production. Those transgenic flies induced less courtship behavior from normal males. *Fad2* was not expressed in males of *D. melanogaster* and *D. sechellia* but was involved in the production of 5,9-HD in females of *D. melanogaster* African strains. Both sexes in *D. simulans* did not express *Fad2*. It played an important role in courtship behavior of *D. melanogaster* as its expression is responsible of female-specieifc CHs (Chertemps et al. 2006; Legendre et al. 2008).

Some researchers have proposed that *desaturases* in *D. melanogaster* may experience similar evolution progress with moth *desaturases* (Knipple et al. 2002). Nevertheless, the evolution of *desaturase* genes in *Drosophila* species still remains unclear due to the lack of information of *desaturases* in most *Drosophila* species. In this chapter, I aimed to study the evolution of *desaturase* genes in *Drosophila,* with special

emphasis on *Fad2*. My first objective is to see if the evolution of *desaturases* in *Drosophila* species also follows the birth-and-death model demonstrated in moths. I searched for the *desaturase* homologs in 12 available *Drosophila* genomes and reconstructed the phylogeny of *desaturase* genes.

My second objective is to understand how *Fad2* arosed by gene duplication. Because *Fad2* is an intron-less gene, it is most likely the product of retrotransposition (Bai et al. 2007). A gene duplicated by retrotransposition has the following four characteristics: lost of introns, lost of regulatory regions, a poly A tail and short flanking direct repeats (Rogers 1985). We will examine *Fad2* homologs in 12 *Drosophila* species for those characteristics and infer the origin of *Fad2* from our results.

The last objective of this chapter is to study the functional perspective of *Fad2* expression in *Drosophila* species. To do so, the expression patterns of *Fad2* homologs were examined and compared with available CHs data (Fig 2.1). Based on the *Fad2* function in *D. melanogaster*, it is responsible for female-specific diene production with unsaturated substrates. If the functions of *Fad2* homologs are conserved in other *Drosophila* species, I would expect to link the expression of *Fad2* to the CH phenotypes that female should possess female-specific diene product.

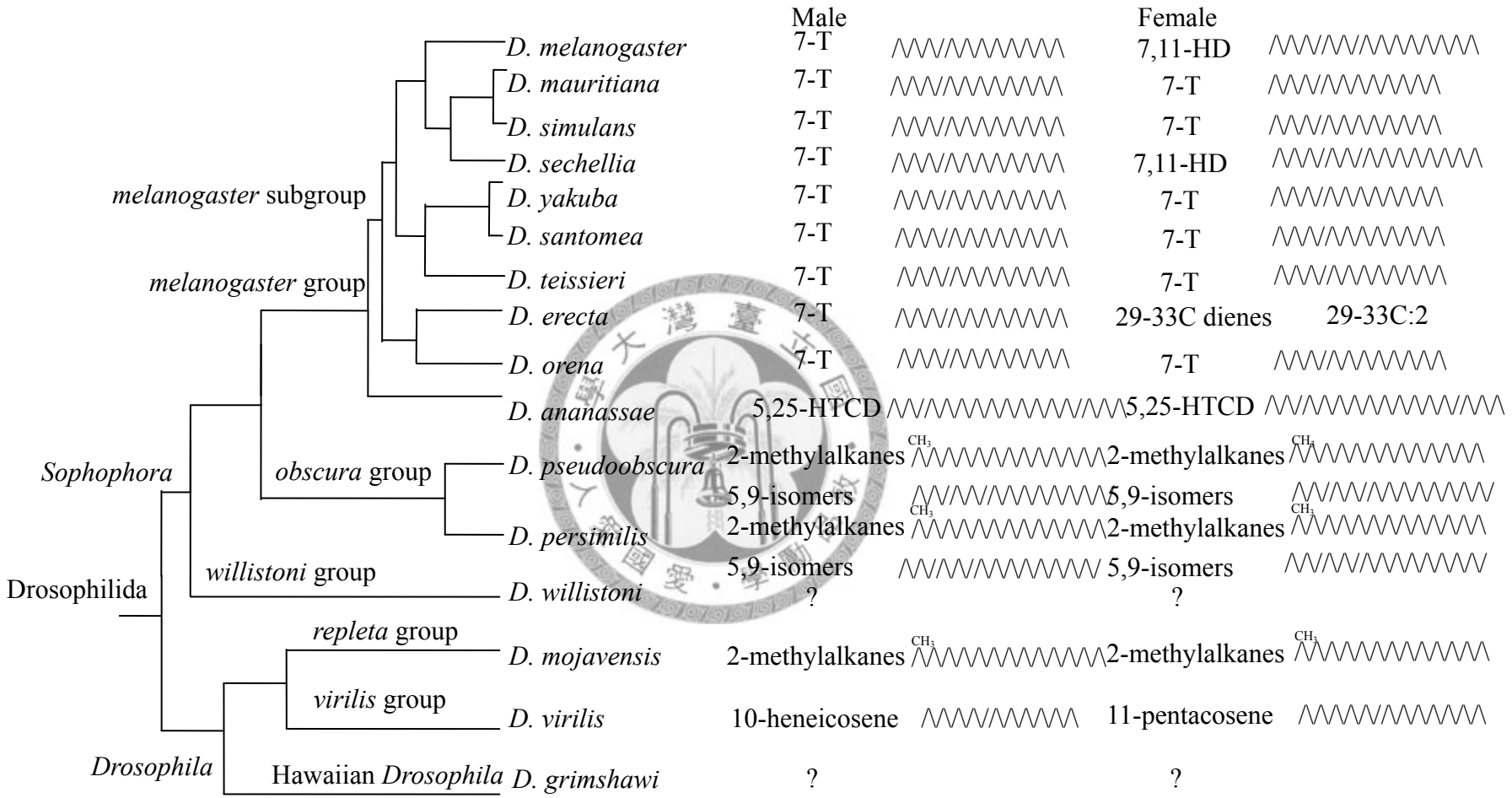


Fig 2.1. Comparison of the main CHs in Drosophilidae (Ferveur 2005; Wicker-Thomas 2007). 2-Methylalkanes including methyl branched alkanes, range from 25C-31C, 5,9-Isomers range from 25-29C.

2.2.

Materials and Methods

Fad2 homologs in 12 *Drosophila* genomes

The nucleotide and amino acid sequences of *Fad2* and other desaturase genes (*desat1*, *desat2*, *CG8620*, *CG9743*, *CG9747* and *CG15531*) were obtained from *D. melanogaster* genome sequences (R5.7) (Kaminker et al. 2002; Misra et al. 2002; Celinker et al. 2002) and performed BLAST in database Genome Assembly (NT), Annotated genes (NT), and Annotated proteins (AA). The BLAST results of other 11 *Drosophila* genomes were from *D. pseudoobscura* genome (R2.2) (Stark et al. 2007; Richards et al. 2005; Clark et al. 2005) and comparative analyses freeze 1 (CAF1) of *D. simulans, D. sechellia, D. yakuba, D. erecta, D. ananassae, D. persimilis, D. willistoni, D. mojavensis, D. virilis* and *D. grimshawi* genomes (Stark et al. 2007; Richards et al. 2005). I searched 500 bp of *Fad2* homologs flanking regions at the syntenic positions for retrotransposition characteristics. I also searched repeat sequences by CENSOR (Kohany et al. 2006) and used Compare and Dotplot program in Genetic Computer Group package (GCG) to compare those sequence regions to verify if there was *Fad2* homologs in the other 11 species. Compare program compares two proteins or nucleic acid sequences and creates a file of the points of similarity between them and the output file will be made a dot-plot by DotPlot (http://www.hku.edu/bruhk/gcgdoc/gcgmanual.html). To

identify their chromosome locations surveyed at least five genes in the flanking regions of annotated *Fad2* homologs.

Sequence analyses

Sequences downloaded from FlyBase were analyzed by multiple sequences alignment in MegAlign program of DNASTAR (Lasergene software package). The genealogy amino acid sequences of seven *desaturase* genes were reconstructed by the Molecular Evolutionary Genetic Analysis 4.0 (MEGA 4.0) software using the neighbor-joining (Saitou and Nei 1987) with Poission correction model and maximam parsimony methods (Fitch 1971). Both of them were performed with 1000 bootstrap replications. I used MrBayes (Ronquist and Huelsenbeck 2003) to infer the phylogeny with two variable-rate models, GTR and F80, and running six Markov chains to approximate posterior probabilities with 1500000 generations. The GTR model allows the stationary state frequencies and substitution rates of amino acids to be different and F80 model also allows the stationary state frequencies but assume the same substitution rates. I used *infertile crescent* (*ifc*) from *D. melanogaster* as outgroup. The pairwise d_N/d_S ratios were estimated by PAML4 (Yang 2007) (runmode= -2, CodonFreq = 2 in codeml.ctl).

Putative transcription factor binding sites in the 5' flanking regions of *Fad2* homologs

were analyzed by PATCH program in BIOBASE (http://www.gene-regulation.com/cgi-bin/pub/programs/patch/bin/patch.cgi?). We only presented the results that were perfectly matched.

Fly stocks

Fourteen strains from 11 *Drosophila* species were used (Table 2.1). Four to five-day-old virgin flies were used for expression analysis. All flies were kept on cornmeal medium at 22°C with a 12-hr light/dark cycle. I added *D. santomea* and *D. teissieri* to RT-PCR study because they and *D. simulans* are all belong to *D. melanogaster* subgroup and share similar CHs phenotype.

Table 2.1. The list of flies used for expression analysis.

Species	Strain
D. melanogaster	Canton-S
D. simulans	W81 (284.19)
D. sechellia	TSC# 14021-0248.25, S9
D. yakuba	TSC# 14021-0261.01, CY19
D. santomea	LAGO 1482
D. teissieri	CT01, GT69w
D. erecta	E220-5 (286.1)
D. ananassae	TSC# 14024-0371.13
D. pseudoobscura	TSC# 14011-0121.94
D. persimilis	TSC# 14011-0111.49
D. willistoni	TSC# 14030-0811.24

RNA preparation and reverse-transcription polymerase chain reaction (RT-PCR)

Total RNA were extracted from five virgin flies of each sex from each strains by using TRIZOL® Reagent (Cal. No. 15596-026, Invitrogen), following the manufacturer's instructions. RNA was dissolved in DEPC water and stored at -70°C. The dissolved RNA was treated by DNase I, Deoxyribonuclease I, Amplification Grade (Cal. No. 18068-015, Invitrogen).

The first-strand cDNA synthesis reactions were performed by Superscript™ III First-Strand Synthesis System for RT-PCR (Cal. No. 18080-51, Invitrogen). Gene specific primers (Table S1) were used to amplify target cDNA sequences by PCR. Amplification was performed in a 20 μl reaction volume. Reactions were run with 30 cycles of 30 s at 95°C for denaturation, 30 s at 55 and 60°C for annealing, 1 min at 72°C for extension, and the final extension at 72°C for 10 min. The PCR amplification products were analyzed by electrophoresis in 1% agarose gels.

2.3.

Results

Desaturase genes in *Drosophila*

Twelve *Drosophila* species contained one homolog for each *desaturase* gene with some exceptions (Table S2): *desat2* was absent in *D. erecta* and multiple *Fad2* homologs were found in *D. ananassae, D. pseudoobscura, D. persimilis* and *D. willistoni*. The phylogeny of seven *desaturase* genes in 12 *Drosophila* species (Fig 2.2A) showed that each *desaturase* genes formed a monophyletic group. The lineage of *desat1* and *desat2* were closely-related and *Fad2* formed another distinct but closely-related branch. For the other four *desaturase* genes (*CG8630*, *CG9743*, *CG9747* and *CG15531*), *CG8630* was the closest one to the lineage of *desat1*, *desat2* and *Fad2*. *CG9743* and *CG9747* were closely-related and joined together into an individual branch. *CG15531* was the most distinct one from other six *desaturase* genes. The Bayesian tree with F80 model showed similar results with neighbor-join and maximum parsimony tree (Fig 2.2A). However, the Bayesian tree with GTR model, *Fad2* cluster was closely-related to the ourgoup (Fig 2.2B).

There was no annotated *desat2* in *D. erecta* and the Dotplot result (Fig 2.3) showed that only a small fragment of *desat2* sequence was found in the syntenic position. There were three annotated *Fad2* homologs in *D. ananassae, D. pseudoobscura, D. persimilis* and two in

D. willistoni. In *D. ananassae*, all of them (*GF16174, GF24026,* and *GF18504*) were not in the syntenic position. *GF16174* and *GF18504* were in the Muller element B and *GF24026* was on the Muller element C, homologous to the right and left arm of the third chromosome in *D. melanogaster*. In *D. pseudoobscura* and *D. persimilis*, there were also three *Fad2* homologs. Two *Fad2* homologs (*GA20691* in *D. pseudoobscura* and *GL15669* in *D. persimilis*) were in the syntenic position and the others (*GA27148* and *GA27452* in *D. pseudoobscura*, *GL23117* and *GL22317* in *D. persimilis*) were on the Muller element E, homologous to third chromosome right arm in *D. melanogaster*. The two *Fad2* homologs (*GK17186* and *GK11373*) in *D. willistoni* were both on the Muller element D, homologous to third chromosome left arm in *D. melanogaster*. However, they were not in the syntenic position. The precise chromosome location of *GK11373* was hard to identify due to the variation of flanking genes. The flanking genes of *GK11373* were on the same chromosome arm but their sequences were discontinued. The similar pattern is also observed in *GF18504* of *D. ananassae.* There were no annotated *desaturase* genes in the chromosome locations of those additional *Fad2* homologs in other *Drosophila* species. There was no annotated *Fad2* homolog in *Drosophila* subgenus including *D. mojavensis*, *D. virilis* and *D. grimshawi*. The Dotplot results also revealed that there was no missing annotated *Fad2* homolog in the syntenic position of *D. mojavensis*, *D. virilis* and *D. grimshawi*.

Retrotranspositions possessed four molecular features: the lost of regulatory regions, no introns, poly-A tail and flanking by short direct repeats. The CENSOR results showed poly-A sites were found out of cases although no repeat sequence within 500 bp flanking regions of the syntenic *Fad2* homologs could be identified.

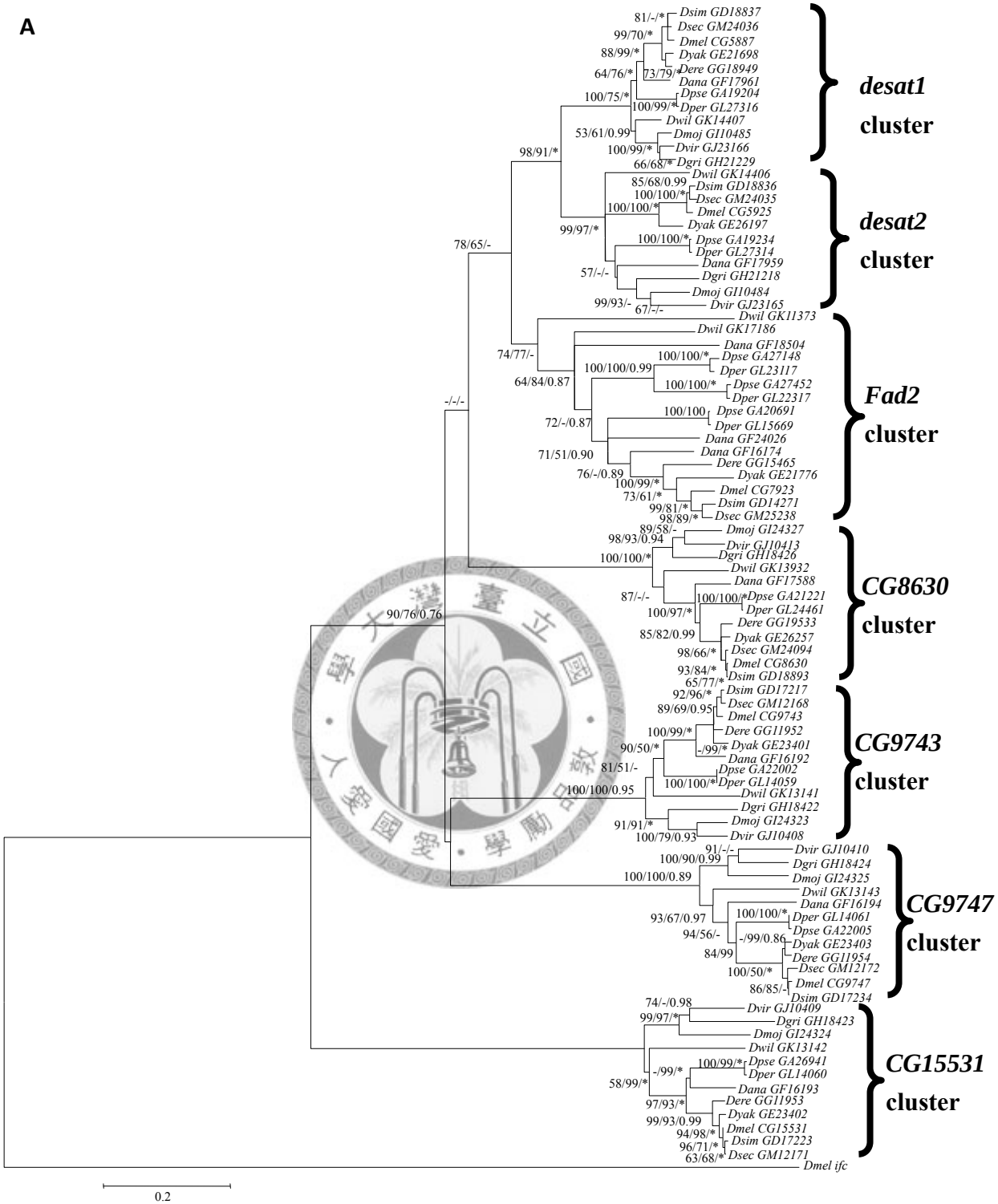


Fig 2.2. The phylogenic tree of seven *desaturase* genes from 12 *Drosophila* species. (A) The numbers above the nodes are bootstrap values of Neighbor-Joining, Maximum Parsimony method (higher than 50 are shown) and posterior probability of Bayesian method with F81 model (higher than 0.85 are shown). (B) The Bayesian tree with GTR model. The numbers above the nodes are posterior probability (higher than 0.85 are shown). *= 1.00

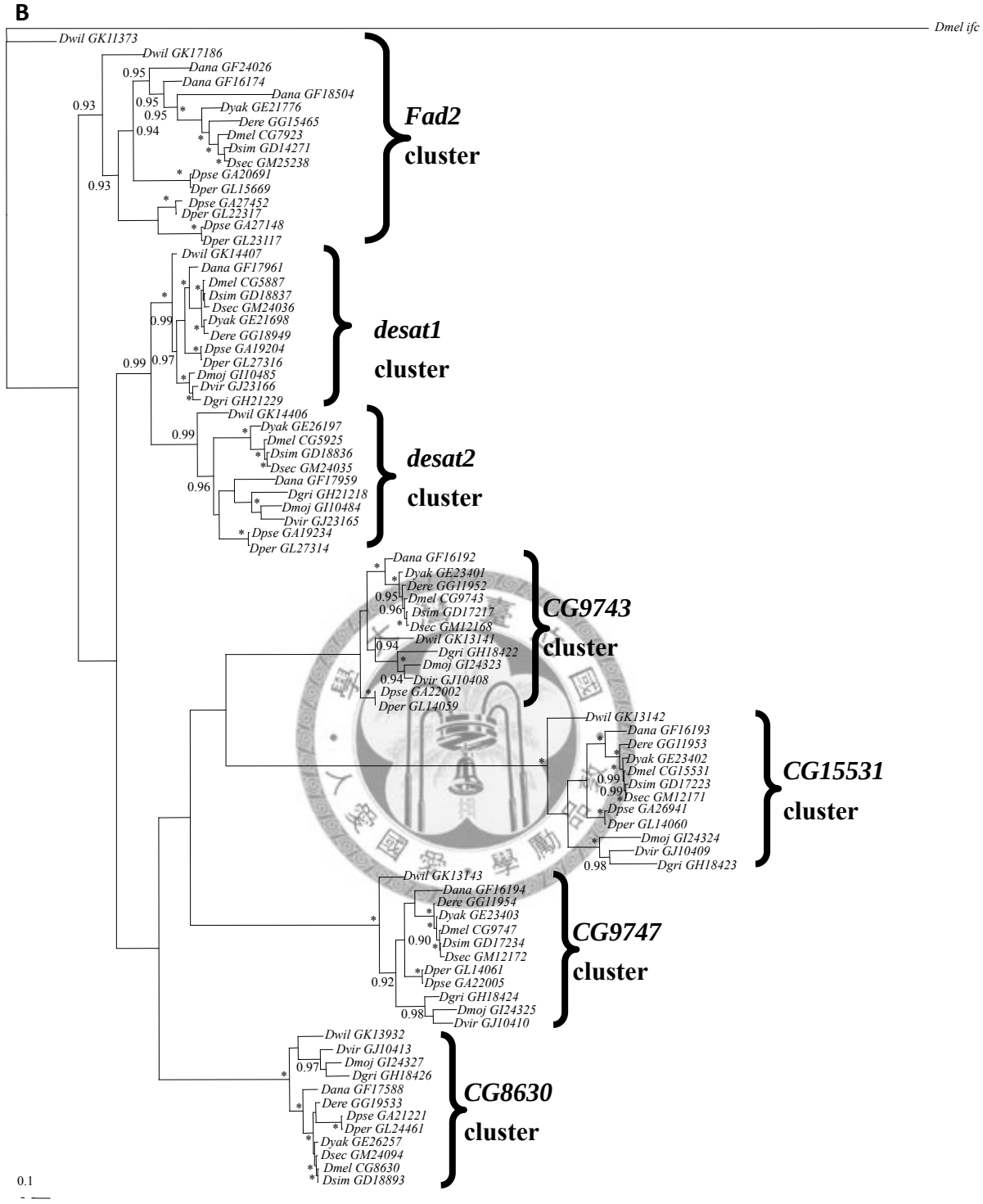


Fig 2.2 (continued)

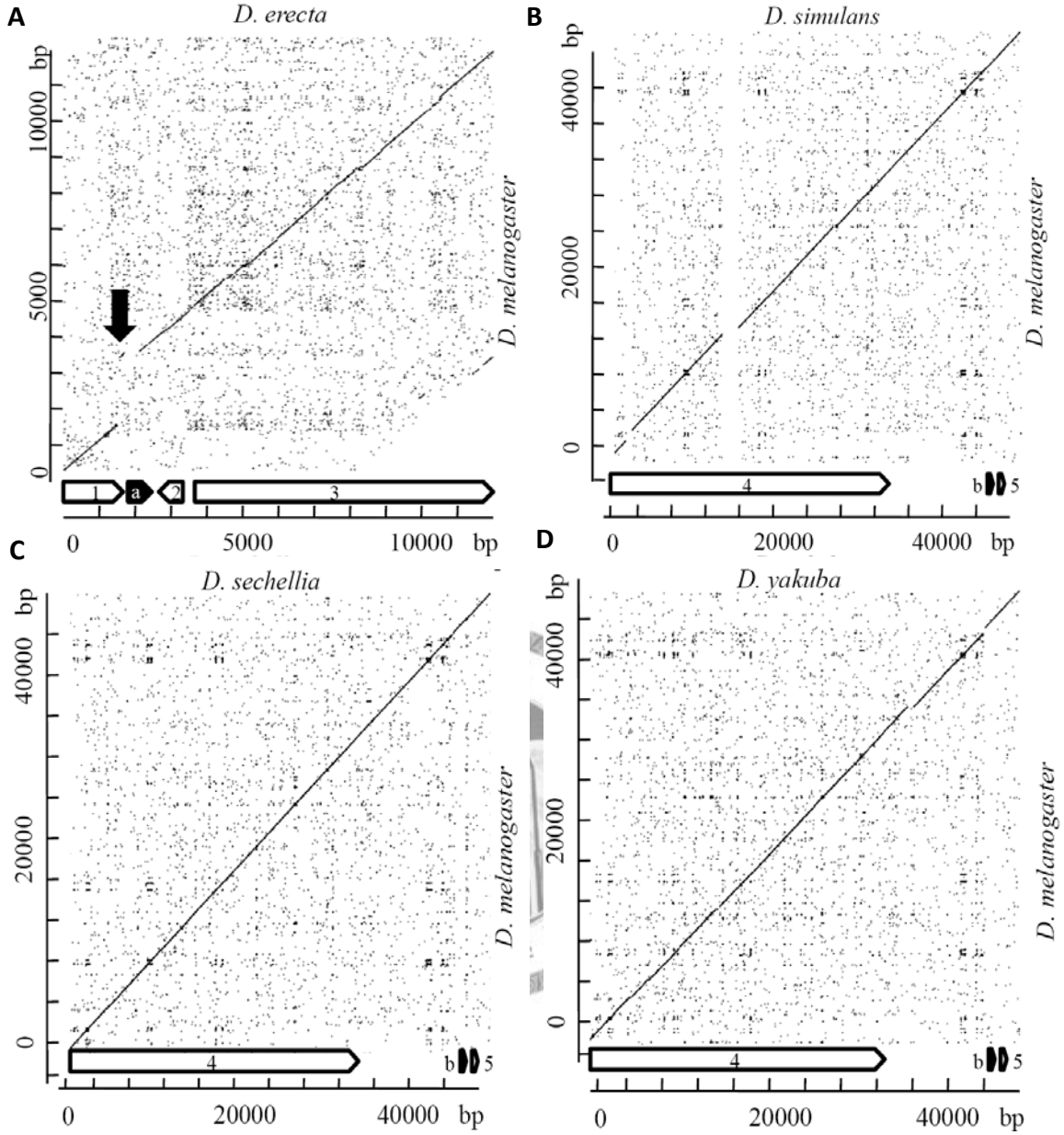


Fig 2.3. The Dotplot of *desaturase* genes. A: The comparing of *desat2* in *D. melanogaster* with *D. erecta*. B-L: The comparing of *Fad2* in *D.melanogaster* with other 11 available *Drosophila* genome. The comparing included the 5' and 3' flanking genes. The black arrow points the residues of *desat2* in *D. erecta*. 1: *CG5844*, 2: *CG17207*, 3: *desat1*, 4: *klu*, 5: *CG32079*, 6: *CG6327*, a: *desat2*, b: *Fad2*.

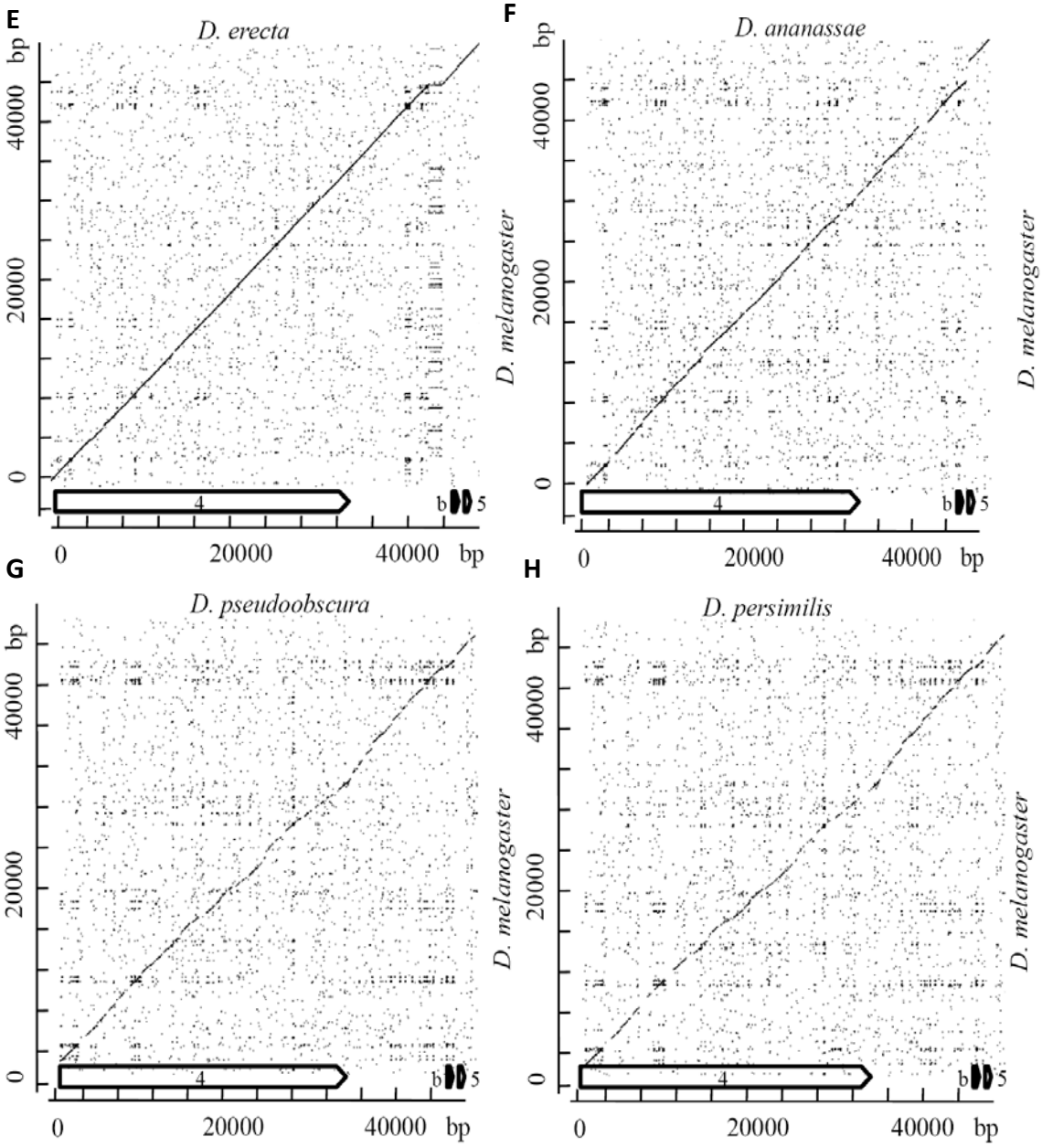


Fig 2.3 (continued)

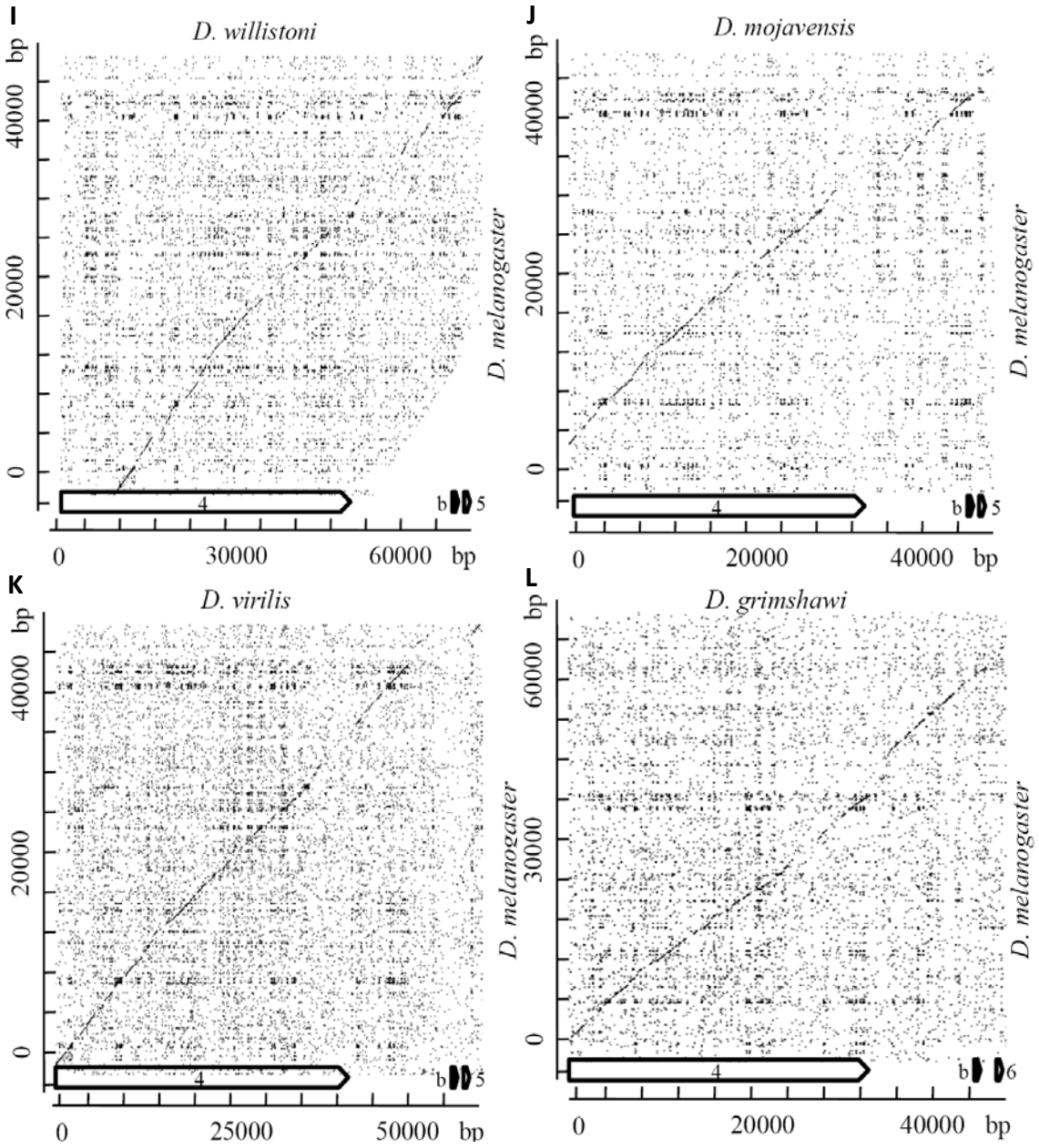


Fig 2.3 (continued)

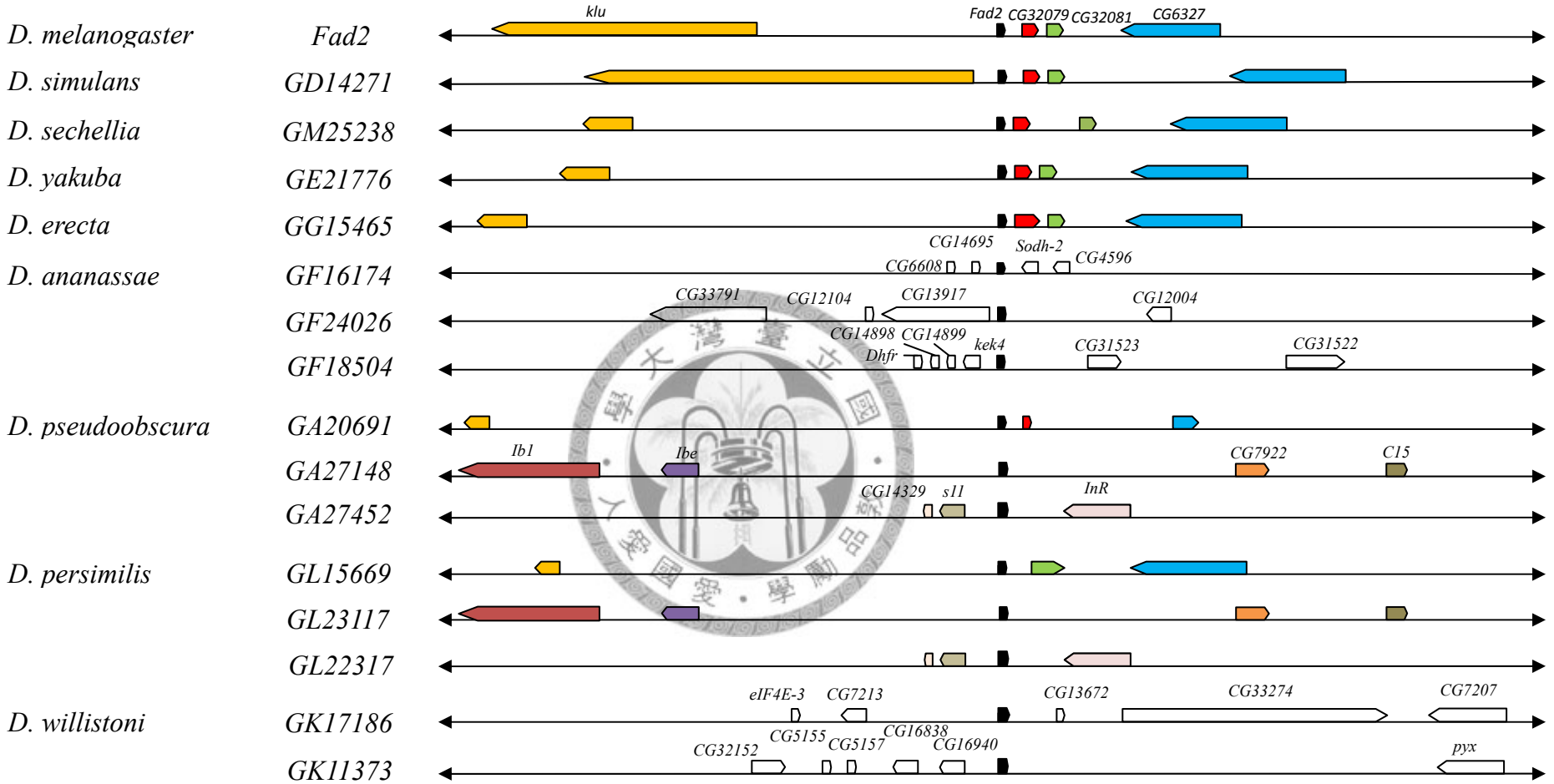


Fig 2.4. Mapping of *Fad2* homologs in *Drosophila*. The homologs genes were searched by FlyBase BLAST tool. Multiple homologs are found in *D. ananassae*, *D. pseudoobscura*, *D. persimilis* and *D. willistoni*. All symbols are from *D. melanogaster* homologs.

Fad2* in *Drosophila

The analysis of *Fad2* homologs showed there were no degeneration occurred except in *GE21776* (*D. yakuba*). There was a 10 bp deletion in the coding sequence and an annotated 32 bp intron that would result in a shorter transcript (1020 bp) in *D. yakuba*. The first 300 bp of coding regions were the most divergent in *D. pseudoobscura*, *D. persimilis*, and *D. willistoni.* The *Fad2* homologs in *D. pseudoobscura*, *D. persimilis*, and *D. willistoni* were longer than other species (1089 bp in *GA27148* of *D. pseudoobscura* and *GL23117* of *D. persimilis*; 1086 bp in *GA27452* of *D. pseudoobscura* and *GL22317* of *D. persimilis*; 1092 bp in *GK17186* and *GK11373* of *D. willistoni*) because their 5' coding regions had additional sequences. The length of other *Fad2* in other homologs except *GE21776* in *D. yakuba* was 1065 bp. The annotated four transmembrane domains and three histidine boxes were highly conserved among sixteen *Fad2* homologs (Fig 2.5). The integrity of hisitidine box was very important because it was essential for desaturase catalytic activity (Knipple et al. 2002). The pairwise d_N/d_S ratios of all *Fad2* homologs were less than one (Fig 2.6).

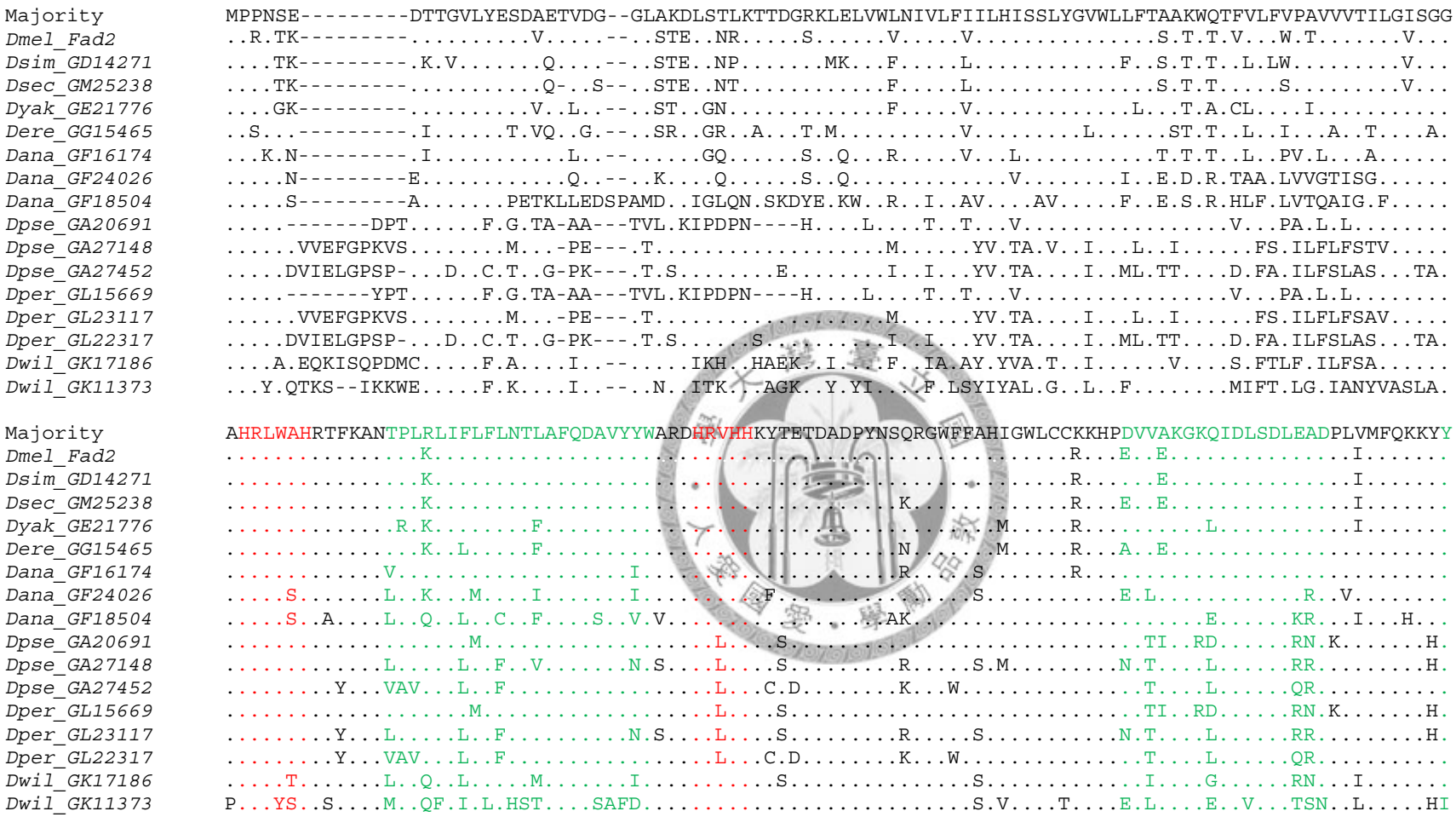


Fig 2.5. The alignment of *Fad2* homolog amino acid sequences in *Drosophila*. Three histidine boxes are shown in red and four annotated transmembrane domains are labelled in green.

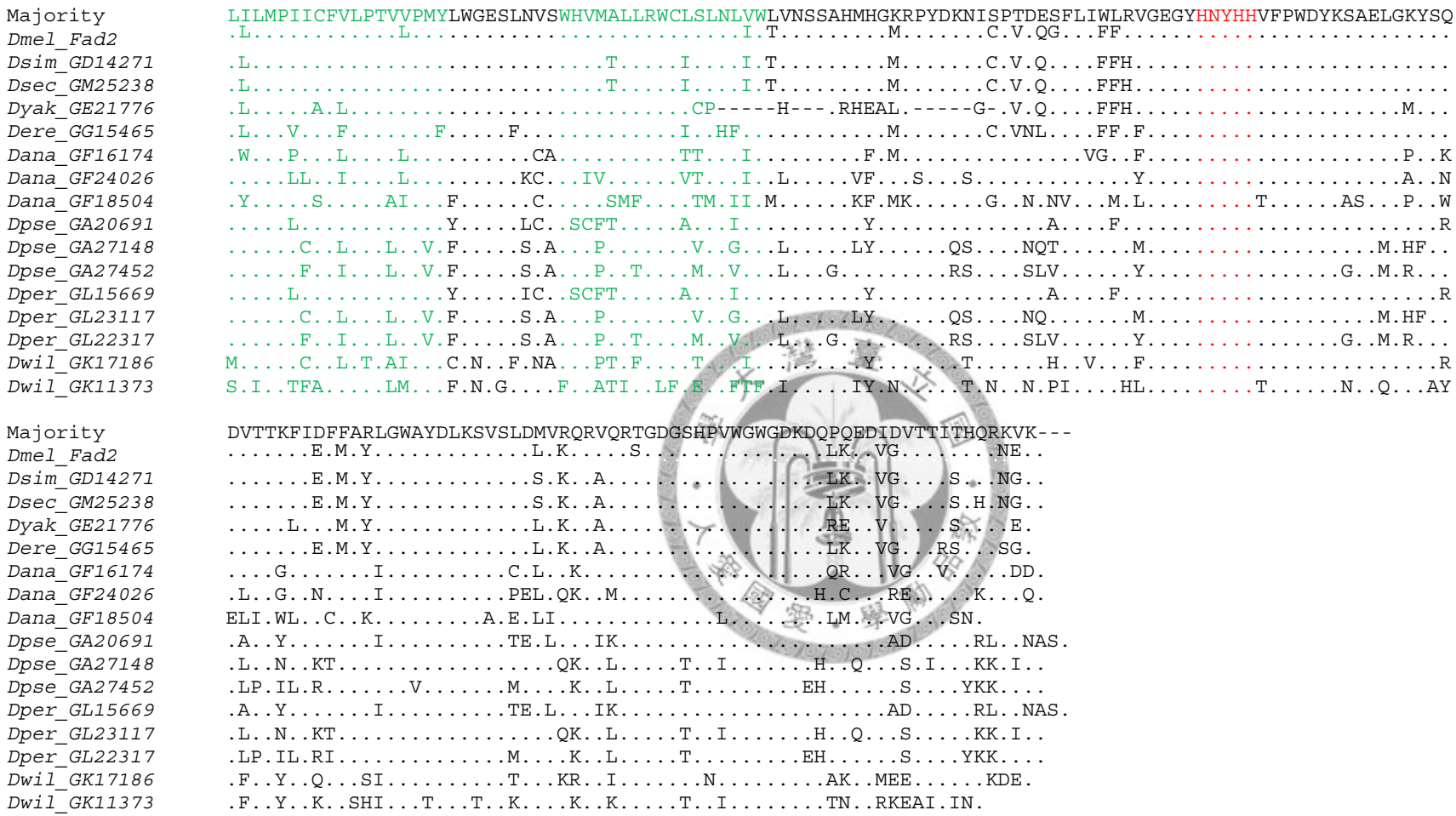


Fig 2.5 (continued)

d_N/d_S

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	1		0.035/0.149	0.026/0.177	0.072/0.598	0.069/0.354	0.121/3.072	0.168/2.382	0.227/2.428	0.169/3.401	0.193/4.673	0.205/73.903	0.167/4.102	0.188/6.717	0.215/74.098	0.214/2.251	0.315/2.537
	2	0.237		0.015/0.063	0.074/0.525	0.067/0.338	0.127/2.823	0.184/2.832	0.235/2.044	0.179/3.358	0.202/5.407	0.221/4.251	0.177/4.056	0.196/7.691	0.220/4.535	0.213/67.611	0.314/2.742
	3	0.148	0.236		0.071/0.537	0.064/0.349	0.123/3.949	0.178/1.898	0.236/1.896	0.166/6.668	0.195/4.347	0.214/5.702	0.164/9.327	0.190/5.677	0.214/5.201	0.210/67.393	0.310/2.872
	4	0.121	0.140	0.132		0.089/0.473	0.143/2.454	0.189/2.588	0.245/2.616	0.201/2.264	0.188/10.063	0.215/77.048	0.199/2.340	0.187/82.666	0.215/77.078	0.233/69.049	0.328/5.477
	5	0.175	0.198	0.182	0.189		0.125/2.297	0.182/3.398	0.238/3.319	0.198/2.908	0.196/4.206	0.208/27.258	0.196/3.218	0.195/5.208	0.206/30.657	0.228/66.044	0.335/2.275
	6	0.039	0.045	0.031	0.058	0.054		0.110/81.501	0.186/7.671	0.161/48.292	0.170/22.261	0.196/16.437	0.160/89.546	0.162/72.787	0.196/17.026	0.185/68.934	0.309/63.937
	7	0.070	0.100	0.094	0.073	0.054	0.001		0.210/1.765	0.175/3.371	0.182/3.159	0.204/2.938	0.174/2.898	0.180/3.217	0.203/3.706	0.193/73.355	0.266/69.291
ω	8	0.093	0.115	0.124	0.094	0.103	0.024	0.119		0.239/86.344	0.249/77.418	0.261/74.283	0.238/85.795	0.252/76.744	0.264/74.746	0.249/69.199	0.332/67.251
	9	0.050	0.053	0.025	0.089	0.068	0.003	0.052	0.003		0.185/54.209	0.206/97.259	0.001/0.066	0.184/76.654	0.205/97.004	0.186/6.454	0.303/6.480
	10	0.041	0.037	0.045	0.019	0.047	0.008	0.058	0.003	0.003		0.105/0.450	0.184/87.724	0.009/0.062	0.102/0.446	0.179/8.095	0.265/3.873
	11	0.003	0.052	0.038	0.003	0.008	0.012	0.069	0.004	0.002	0.233		0.205/96.492	0.097/0.448	0.005/0.036	0.209/69.607	0.311/5.162
	12	0.041	0.044	0.018	0.085	0.061	0.002	0.060	0.003	0.018	0.002	0.002		0.183/106.579	0.203/96.215	0.184/8.148	0.301/6.164
	13	0.028	0.026	0.034	0.002	0.039	0.002	0.056	0.003	0.002	0.140	0.218	0.002		0.094/0.444	0.174/7.209	0.272/3.810
	14	0.003	0.049	0.041	0.003	0.007	0.011	0.055	0.004	0.002	0.228	0.152	0.002	0.213		0.211/69.714	0.313/5.072
	15	0.095	0.003	0.003	0.003	0.003	0.003	0.003	0.004	0.029	0.022	0.003	0.023	0.024	0.003		0.254/1.851
	16	0.124	0.114	0.108	0.060	0.147	0.005	0.004	0.005	0.047	0.068	0.060	0.049	0.071	0.062	0.137	

Fig 2.6. The pairwised d_N/d_S of DNA sequences among 16 *Fad2* homologs. 1: *Dmel_Fad2*, 2: *Dsim_GD14271*, 3: *Dsec_GM25238*, 4: *Dyak_GE21776*, 5: *Dere_GG15465*, 6: *Dana_GF16174*, 7: *Dana_GF24026*, 8: *Dana_GF18504*, 9: *Dpse_GA20691*, 10: *Dpse_GA27148*, 11: *Dpse_GA27452*, 12: *Dper_GL15669*, 13: *Dper_GL23117*, 14: *Dper_GL22317*, 15: *Dwil_GK17186*, 16: *Dwil_GK11373*.

Fad2 expression patterns among *Drosophila*

To investigate the *Fad2* expression patterns among *Drosophila* species, RT-PCR were performed. In *D. melanogaster* subgroup (Fig 2.7). *Fad2* was expressed in females of *D. melanogaster* (Canton S) and *D. erecta* and both sexes of *D. sechellia* (TSC# 14021-0248.25 and S9). The expression of *Fad2* in *D. sechellia* seemed different between strains: males of TSC# 14021-0248.25 had a higher expression level of *Fad2* than S9 males. There was no *Fad2* expression in both sexes of *D. santomea* (LAGO 1482) and two *D. yakuba* strains (TSC# 14021-0261.01 and CY19), which possessed sexually monomorphic CHs. There was no detectable expression for both sexes of *D. santomea*, *D. simulans*, *D. teissieri* and *D. yakuba*.

In *D. ananassae*, both sexes had expressions of *GF24026* and *GF18504*. There was no detectable expression of *GF16174* in both sexes though the expression of *Act5c*, a RNA quality control, was not very strong. It was possible that the level of *GF16174* expression was too low to be detectable by RT-PCR.

In *D. pseudoobscura* and *D. persimilis*, only one *Fad2* homolog, *GA20691* and *GL15669*, respectively, had expressions. Males had a higher expression level than females in *D. pseudoobscura*, whereas both sexes in *D. persimilis* had the same expression level. There was no detectable expression for other *Fad2* homologs of *D. pseudoobscura* (*GA27148* and *GA27452*) and *D. persimilis* (*GL23117* and *GL22317*).

Both two *Fad2* homologs of *D. willistoni* were expressed in both sexes and one of them, *GK11373,* had different expression levels between sexes.

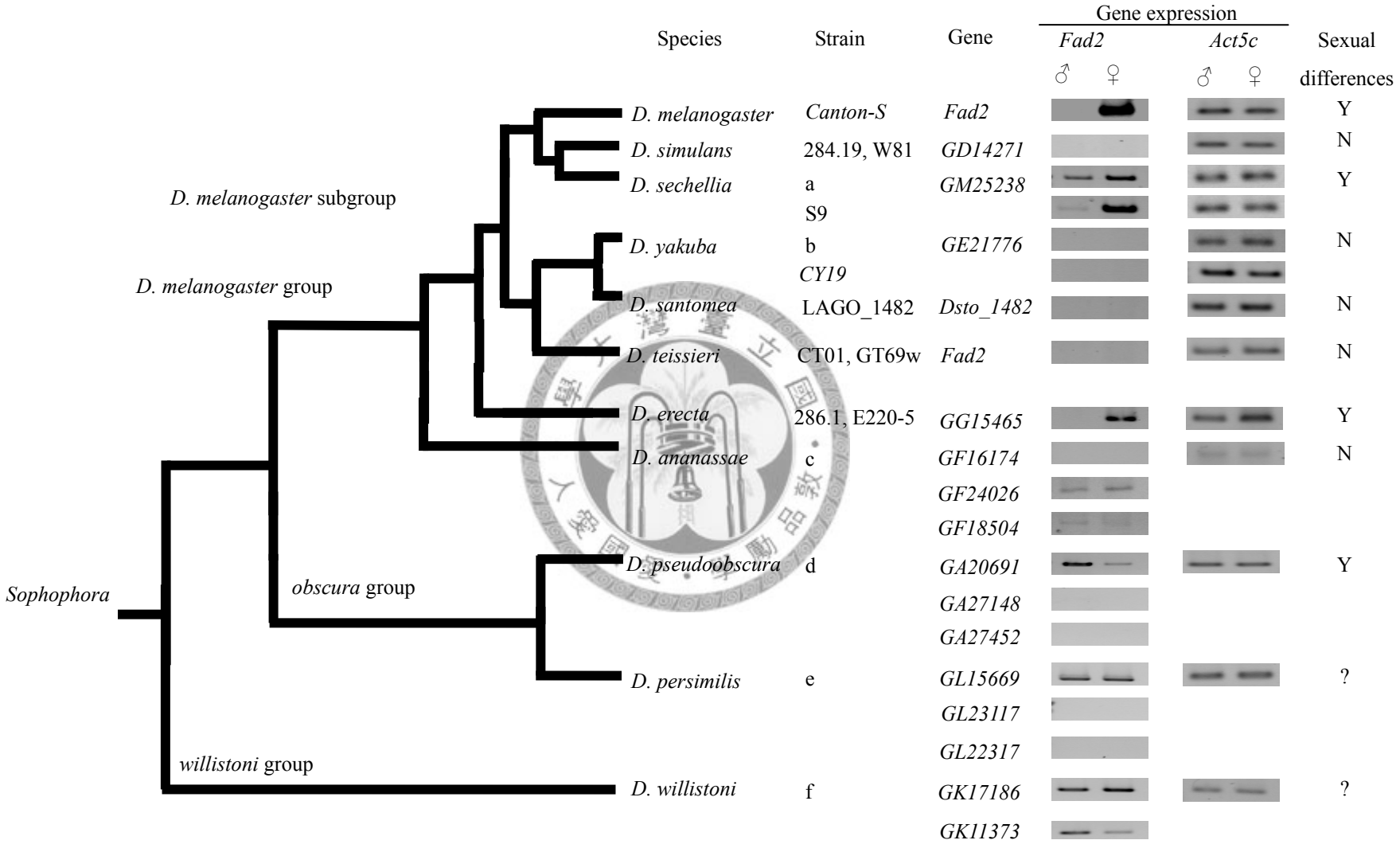


Fig. 2.7. The phylogenetic relationships between *Fad2* homolog expression patterns and sexually dimorphic CHs in *Drosophila* species. (a: TSC#14021-248.25, b: TSC#14021-0231.01, c: TSC#14024-0371.13, d: TSC#14011-0121.94. e: TSC#14011-0111.49, f: TSC#14030-0811.24.)

2.4.

Discussion

The Origin of *Fad2*

The *desaturase* genes in *Drosophila* evolve follow the birth-and-death model. New genes are created by repeated gene duplication, and some duplicated genes are maintained in the genome or become nonfunctional by deleterious mutations (Roelofs and Rooney 2003). *desat2* is deleted in one species (*D. erecta*) and *Fad2* shows the signatures of degeneration (*D. yakuba*) and multiple duplication events (*D. ananassae, D. pseudoobscura*, *D. persimilis* and *D. willistoni*) in different species. I also found that the substitution rates in *Fad2* homologs might be most variable among the seven *desaturase* genes. The Bayesian tree with GTR model was different from the tree reconstructed by F80 model. The evolution of *Fad2* in *Drosophila* must be different from other *desaturase* genes. Previous study had shown that *Fad2* is responsible for female-specific sex pheromone production in *D. melanogaster* and *D. sechellia* (Chertemps et al. 2006). The distinct evolution patterns of *Fad2* in different species reflect the diversity of sex pheromone among *Drosophila* species. It might be a good example of functional shift from metabolism to reproduction in *Drosophila*.

Fad2 is an intron-less gene and is likely to arise through retrotransposition (Bai et al. 2007). We found an A-rich region at the 3' regions of *Fad2* homologs in five species

(*D. melanogaster, D. simulans, D. sechellia, D. yakuba and D. erecta*) though the length were variable and nearly 300 bp away from stop codon but we did not find direct repeats. Since those characteristics might only be retained for a short time after retrotransposition occurring (Long 2001) and *Fad2* showed highly sequence identity with *desat1* and *desat2,* it is strongly suggested that *Fad2* is duplicated by retrotranspositon from *desat1* or *desat2*. My phylogenies all showed that *Fad2* was close but distinct form *desat1* and *desat2* (Fig 2.2A-B). According to previous studies, *Fad2* was duplicated from an ancestral gene that gave rise to *desat1* and *desat2* (Roelofs and Rooney 2003). There are two hypotheses about the origin of *Fad2*: Since *Fad2* homologs are only found in *Sophophora* species, this event is probably happened in the lineage leading to this subgenus; According to the study of *Anopheles gambiae*, it had no *Fad2* orthologs (Roelofs and Rooney 2003). The retrotransposition also is likely to happen before the splitting of *Sophophora* and *Drosophila* subgena and subsequently lost in the latter. The phylogeny of *desaturase* genes revealed that *Fad2* is closely-related to *desat1* and *desat2*. *Fad2* is likely to be duplicated through retrotransposition from *desat1* which give rise to *desat2* by tandom duplication later during the evolution of *Drosophila* (Fig 2.8). In *D. melanogaster* and *D. sechellia*, *Fad2* is responsible for sexually dimorphic CHs. Its function was lost in *D. simulans* and *D. yakuba*. *Fad2* is duplicated in *D. ananassae*, *D. pseudoobscura*, *D. persimilis* and *D.*

willistoni. In *D. ananassae* and *D. willistoni* (Table S2), the original copy could be lost or change to different genetic location by transposition or chromosome rearrangement. *Fad2* in *D. melanogaster* subgroup species (*D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta*) evolved differentially from the species out of this subgroup and they have no additional *Fad2* homologs.

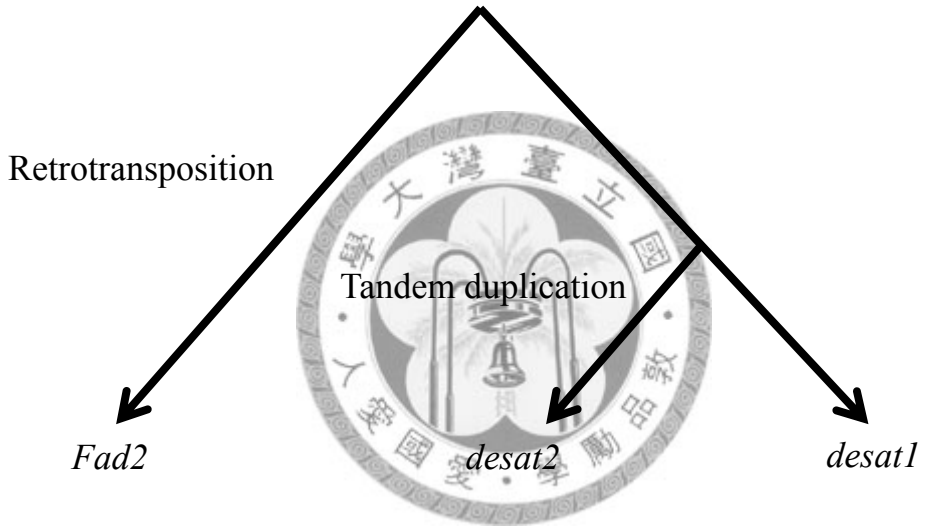


Fig 2.8. The hypothesized origin of three *desaturase* genes in *Drosophila*.

The relationship between *Fad2* and sexually dimorphic CHs in *Drosophila*

There is no correspondence between *Fad2* and sexually dimorphic CHs in *Drosophila* species outside of *D. melanogaster* subgroup. Even without *Fad2* homologs, some *Drosophila* species still have sexually dimorphic CHs, such as *D. virilis*. In nine species of *D. mealanogaster* subgroup, only three of them, including *D. erecta,*

D.melanogaster and *D. sechellia,* had sexually dimorphic CHs. Males of *D. melanogaster* and *D. sechellia* were rich in 7-T and females in 7,11-HD or 5,9-HD. In *D. erecta*, males were rich in 7-T, whereas females had more 29-33C diene. In other species within *D. melanogaster* subgroup, *D. mauritiana*, *D. orena, D. santomea, D. simulans*, *D. teissieri*, and *D. yakuba,* both sexes were rich in 7-T (Antony and Jallon 1982; Jallon 1984; Jallon and David 1987; Cobb and Jallon 1990; Jallon and Wicker-Thomas 2003; Mas and Jallon 2005). *Fad2* is expressed in *D. melanogaster*, *D. sechellia* and *D. erecta* with sexually dimorphic CHs but not in sexually monomorphic CHs species (*D. sumulans, D. yakuba, D. teissieri* and *D. santomea*). The production of diene is related to the sexual isolation among *D. melanogaster*, *D. sechellia* and *D. simulans*, because diene can inhibit *D. simulans* males' courtship behavior but induce them from other two species (Legendre et al. 2008). The expression of *Fad2* strongly correlates with the sexually dimorphic CHs among *D. melanogaster* subgroup species. Besides, all species within *D. melanogaster* subgroup possessed only one *Fad2* homologs.

There are two major findings. First, *Fad2* were expressed in both sexes of *D. sechellia*. This result is different from a previous study that *Fad2* in *D. sechellia* is female-specific expression (Chertemps et al. 2006). Since *D. sechellia* males are unique in *D. melanogaster* subgroup to carry 7,11-HD (Jallon and David 1987), it is reasonable

that *D. sechellia* males also express *Fad2* but the expression level in male is weaker than in female. The difference between this study and Chertemps et al. (2006) are probably due to different expression levels between strains (TSC# 14021-0248.25 and S9) or the different experimental approaches. They performed northern blot to detect the protein product of *Fad2* in *D. sechellia*. The probes that they used were designed from *D. melanogaster Fad2* DNA sequence and resulted in a weaker band in the *D. sechellia* females and no detectable band in males (Chertemps et al. 2006).

Second, we found a strong expression of *Fad2* in females of *D. erecta* that the main CHs in females are 29, 31 and 33 carbon chain length hydrocarbons (Jallon and David 1987). We also do not find *desat2* in *D. erecta.* The CHs biosynthesis pathway in *D. erecta* seems different from other *D. melanogaster* subgroup species. The components of CHs in *D. erecta* females are the most complex among species within *D. melanogaster* subgroup. It composed species-specific long chain (29-33C) dienes and most of them are female-specific (Jallon and David 1987). *Fad2* might be functioned as Δ9 desaturase in the production of some female-specific CHs compounds. The interaction of *desat1* and *Fad2* in *D. erecta* merits further investigation. The role of *Fad2* in the CHs production of *D. erecta* provides a good opportunity to study the species-specific CHs evolution.

In *Drosophila* species outside of the *D. melanogaster* subgroup, the expressions of

Fad2 homologs were detected in both sexes (*GF24026* and *GF18504* in *D. ananassae*, *GA20691* in *D. pseudoobscura*, *GL15669* in *D.persimilis*, *GK17186* and *GK11373* in *D. willistoni*). There are non-expressed *Fad2* homologs (*GF16174* in *D. ananassae*, *GA27148* and *GA27452* in *D. pseudoobscura*, *GL23117* and *GL22317* in *D. persimilis*). The fates of duplicated genes include nonfunctionalization, neofuntionalization, and subfunctionalization including the functional redundancy and expression differentiation (Zhang 2003). One copy could shield the second copy from natural selection and the shielded one may become a pseudogene (Force et al. 1999). However, there is no signature of degeneration in the *Fad2* homologs in *D. ananassae*, *D. pseudoobscura* and *D. persimilis*.

The CHs in *D. pseudoobscura* and *D. persimilis* consist a large amount of 5,9 isomers (5,9-PD, 5,9-HD and 5,9-ND) (Toolson and Kuper-Simbron 1989). *Fad2* was shown to be involved in the production of 7,11-HD and 5,9-ND in *D. melanogaster* (Chertemps et al. 2006; Legendre et al. 2008). *GA20691* and *GL15536,* therefore*,* could be responsible for the production of 5,9 isomers in *D. pseudoobscur*a and *D. persimilis*. *D. pseudoobscura* shows inconsistent sexually dimorphic CHs: males have more 5,9-PD than females, though some results show there was also accompany the decrease of 5,9-HD in males (Toolson and Kuper-Simbron 1989). Our results showed there is likely to be different expressions of *GA20691* between the two sexes in *D.*

pseudoobscura. However, there is no enough information about the relationship between *Fad2* homologs and CHs production in *D. pseudoobscura*.

All the CHs in *Drosophila* species with functional *Fad2* orthologs possess a large amount of 7,11-isomers and 5,9-isomers (*D. melanogaster, D. sechellia, D. pseudoobscura* and *D. persimilis*), whereas there are distinct diene products accumulated in *Drosophila* species with only *Fad2* paralogs (*D. ananassae* and *D. willistoni*). A lot of long chain dienes (29C-35C) existing in the CHs of *D. ananassae* and *D. willistoni* (Chuan-Chan Wang, personal communication) and the second double bond might not be added by Fad2 desaturase because those long chain dienes including 5,25-$C_{31:2}$, 8,24-$C_{33:2}$, 9,25-$C_{35:2}$, etc. It is obviously that those *Fad2* paralogs do not function as *Fad2* in *D. melanogaster*. Those *Fad2* paralogs are probably evolved diversely to orthologs, and they might play different roles in the CHs production. The function of *Fad2* ortholog in *D. erecta* is still unclear because the lack of detailed study about the female-specific CHs.

Since I only tested gene expression in adult flies, those non-expressed *Fad2* homologs are probably expressed in other *Drosophila* life stages. The main function of CHs is to prevent the desiccation (Edney 1967), which is necessary in all life stages of insects. In the analysis of 5' regions (Fig S1, Table S3), I found several transcription factor binding sites that were involved in embryo or larva development like Twi, E74,

LEF-1, RXR-alpha, EcR, Zeste, DREF and Cad, whose functions were intercorrelated. For example, Zeste can bind to the E74 promoter. These sites overlap with three GAGA-binding sites. Those overlapped binding are cis-transcriptional control elements (Thummel 1989). The DRE (DNA replication-related element) and DREF (DNA replication-related element factor) can affect the expression of TATA box binding protein that regulates gene expression (Choi et al. 2002). Twi, E74, RXR-alpha and EcR are induced by the moulting pheromone, ecdysone, and in turn, regulate the expressions other genes (Nuesslein-Volhard, Wieschaus, and Kluding 1984; Thummel 1989; Thomas, Stunnenberg, and Stewart 1993). One of my important finding is that we found sex determination factor binding site (DSXF, DSXM) were presented in all 5' regions of most *Fad2* homologs except in *GA27452* (*D. pseudoobscura*), *GL22317* (*D. persimilis*), *GK17186* and *GK17313* (*D. willistoni*). The exhibition of sexually dimorphism is dependent on the activation of sex determination genes. In *Drosophila*, the X chromosome/autosome (X:A) ratio will influence the activity of *Sex-lethal* (*Sxl*), a sex determination gene. If the ratio is equal or more than one (3X: 2A or 2X: 2A), *Sxl* would be activated and the individual is being a female otherwise it will be a male. *Sxl* also regulates the expression of *transformer* (*tra*) the female specific gene (Cline 1993). *tra* and *tra2* are known to activate *doublesex* (*dsx*) and produce the female doublesex protein (DSXF). The male doublesex protein (DSXM) would be expressed when *tra* is

suppressed (Baker et al. 1989; Steinmann-Zwicky, Amrein, and Nothiger 1990). The expressions of *Fad2* homologs are not constrained in females or males of *D. ananassae*, *D. pseudoobscura*, *D. persimilis* and *D. willistoni* even though there were different expression levels between sexes. Further studies are needed to test whether those sites affect the expression of *Fad2* homologs by comparing the differences of 5' regions between sexes.

The binding of DSXF or DSXM would cause the sex-biased expression of genes (Coschigano and Wensink 1993). However, the annotated DSXF and DSXM binding site results are not comsistant with my RT-PCR results that only two *Fad2* homologs (*Fad2* in *D. melanogaster* and *GG15465* in *D. erecta*) show female-specific expression. According to the study of *yolk protein* (*Yp*) gene, the sex-specific expression was regulated by dsxA which can connect the target gene to the sex determination pathway (An and Wensink 1995). Those *Fad2* homologs with DSXF and DSXM binding site might lack the regulation of dsxA resulting their expression patterns were not sex-specific. However, I did not found possible dsxA binding site in the 5' flanking region of all *Fad2* homologs. Perhaps the dsxA binding site is out of the first 1000 bp in the 5' flanking region.

Since *Fad2* is not expressed in sexually monomorphic species within *D. melanogaster* subgroup. The pseudogenization might have happened in *Fad2* of those

species. Is the annotated intron and deletion in *Fad2,* as in *D. yakuba* (*GE21776*) a signature of degeneration? To answer the question, I shall search for the signature of degeneration in sexually monomorphic CHs species within *D. melanogaster* subgroup to explore the evolution of *Fad2* among them.

Chapter 3

The pseudogenization of *Fad2* in *D. melanogaster* subgroup

3.1.

Introduction

Gene duplication can be created by unequal crossing over, retrotransposition and chromosomal duplication. Unequal crossing over results in tandem duplication and the duplicated fragment can contain part of gene, entire gene or several genes. Retrotransposition creates an intron-less duplicated gene. A messenger RNA is retrotranscribed into complementary DNA and inserted into random location in the genome. Chromosomal duplication usually occurs in plants and rarely in animals. This large duplication might be caused by non-disjunction after DNA replication (Zhang 2003).

The duplicated gene will have several evolutionary fates, pseudogenization, subfunctionalization and neofunctionalization. Pseudogenization means that the duplicated gene becomes nonfunctional because there is no selection acting on it and mutations will then accumulate in this pseudogene. Young pseudogenes can be identified by sequence similarity. For example, in *C. elegans*, the ratio of pseudogenes which discovered by sequence similarity and functional genes in the genome is 1:8 (Harrison, Echols, and Gerstein 2001) and 1:2 in humans (Harrison et al. 2002).

Subfunctionalization means that the duplicated gene retains part of its original function. For example, in zebrafish, *engrailed-1* is expressed in the pectoral appendage bud, whereas *engrailed-1b* is expressed in the hindbrain and spinal cord. They are generated by a chromosomal segmental duplication and their ortholog in mouse is expressed in both parts (Force et al. 1999). The duplicated gene can evolve functional specialized if it functions better than the parent gene (Jensen 1976; Orgel 1977; Hughes 1994; Nowak et al. 1997). Sometimes, the duplicated gene evolves new or related function after duplication. However, it needs several amino acid substitution to achieve it (Zhang, Rosenberg, and Nei 1998). For example, duplicated hemoglobin genes in humans are expressed at different developmental stages (reviewed in Ginder, Gnanapragasam, and Mian 2008). The duplicated gene may also conserve the original function and the result will be excessive protein product. The two genes will not be maintained stably in the genome unless the extra expression is an advantage to individual (Nowak et al. 1997).

Fad2 is involved in the biosynthesis of sex pheromone in *D. melanogaster*. It is also a gene with female-specific expression in *D. melanogaster* (Chertemps et al. 2006). Genes related to reproduction and with sex-biased expression usually have faster evolution rates and are highly divergent among species. The male-biased genes are frequently under positive selection due to male-male competition (Proschel, Zhang, and Parsch 2006). For example, the male specific accessory gland proteins (*Acps*) in

Drosophila had been showed under positive selection in different species (Tsaur and Wu 1997; Tsaur, Ting, and Wu 1998; Schully and Hellberg 2006; Wagstaff and Begun 2007; Wong et al. 2008). Female-biased genes tend to be under purifying selection because sexual selection is weak in females (Proschel, Zhang, and Parsch 2006). However, the coevolution between males and females might drive some female-biased genes under positive selection. For example, the male fertilization success is determined by an interaction between sperm and female morphology. So the male sperm length occurred as a correlated response to selection on the female seminal receptacle length. (Pitnick, Markow, and Spicer 1999; Miller and Pitnick 2002).

In previous chapter, I had found that the expression of *Fad2* in species within *D. melanogaster* subgroup would cause sexually dimorphic CHs. *Fad2* is only expressed in the sexually dimorphic CHs species of *D. melanogaster* subgroup (*D. erecta*, *D. melanogaster* and *D. sechellia*) but not in the sexually monomorphic ones including *D. santomea, D. simulans, D. teissieri* and *D. yakuba*. Pseudogenization probably happened to non-expressed *Fad2* in the species with sexually monomorphic CHs.

In order to understand whether *Fad2* is degenerated in sexually monomorphic CHs *Drosophila* species within *D. melanogaster* subgroup. I used the population genetics analyses to study *Fad2* DNA sequences in *D. simulans*, *D. teissieri* and *D. yakuba*. The results can help us understand the pseudogenization of *Fad2* in sexually monomorphic

CHs species within *D. melanogaster* subgroup.

3.2.

Materials and Methods

Fly stock and preparation

Twenty *D. simulans* strains collected from France, United States, Australia, Japan, Reunion Island, Madagascar, Kenya and Zimbabwe, 19 *D. yakuba* strains from Cameroon and 16 *D. teissieri* strains received from Manyuan Long, Cameroon and Gabon were used for population analyses (listed in Table S3). All flies were kept on cornmeal medium at 22°C with a 12-hr light/dark cycle.

Fad2 sequences

DNA was extracted from single fly of each strain following modified Gentra Puregene DNA purification manual. PCR amplification with gene-specific primers (Table S1) was performed in a 20 μl reaction volume. Reactions were carried out with 30 cycles of 30 s at 95°C for denaturation, 30 s at 55°C for annealing, 1 min at 72°C for extension, and the final extension at 72°C for 10 min. The PCR amplification products were analyzed by electrophoresis in agarose gels. Sequencing was done by ABI PRISM® 96-capillary 3730 xl DNA Analyzer. The total length of PCR product was 2690 bp including 1141 bp of 5' flanking region, 1068 bp of coding region and 481 bp of 3' flanking region.

Data analyses

Multiple sequence alignments were obtained by using MegAlign program in DNASTAR Lasergene software package (Clewley 1995; 1997; Clewley and Arnold 1997; Burland 2000). Further analyses were performed by DnaSP 4.50.2 (Rozas et al. 2003). I used following tests to test the neutrality of target sequence regions, including flanking regions and coding regions.

1. Tajima'D test (Tajima 1989): it compares θ (a standardized measure of the number of segregating sites) and π (the average proportion of nucleotide differences overall all pairwise comparison of the samples). Under the null neutral model, the result should be no difference (0). A negative result represents excessive low frequency mutations, indicating population size expansion and/or purifying (negative) selection. A positive result means low levels of low and high frequency mutations, indicating a decrease in population size or balancing selection.
2. Fu and Li's tests (Fu and Li 1993) : it contains D and F tests. They are used to distinguish the internal and external mutations on the genealogical tree. The D test statistic is based on the difference between the total number of singletons and the total number of mutations; The F test statistic is based on the differences between the number of singletons and the average number of nucleotide differences between pairs of sequences. Under the null neutral model, the result should be no difference

(0). A negative result represents excessive medium frequency mutations, indicating positive selection or expanding population. A positive result means balancing selection or population bottleneck.

3. McDonald and Kreitman test (McDonald and Kreitman 1991): it is used to compare the ratio of replacement sites to silence sites between species and within species. Under neutrality, the between-species ratio should be the same as the within-species ratio. A significant difference between the two ratios means positive selection might affect their evolution.

4. Fay and Wu's H test (Fay and Wu 2000): it was design to detect incomplete genetic hitchhiking. Under directional selection, the mutations that are tightly linked to the advantageous one will be carried to fixation. Before reaching the fixation, these mutations will be in high frequency, therefore a measure that gives more weight to mutations of larger sizes will be able to capture the signature of incomplete hitchhiking, and its comparism with π can yield significant result under hitchhiking. The significant negative result means the target sequence possessed excessive high-frequency variants and is under positive selection. I perform the test on the website of the Fay's lab (http://www.genetics.wustl.edu/jflab/htest.html).

3.3.

Results

Pseudonization of *Fad2* in *D. simulans*

After surveying *Fad2* sequences of 21 *D. simulans* strains, 25 bp deletion was found in one strain (W75) that resulted a frameshift mutation (Fig 3.3). However, there was no signature of degeneration observed in the other 20 *D. simulans Fad2* sequences.

The nucleotide diversity measured by both π and θ showed the 5' flanking region of *Fad2* sequences in *D. simulans* populations were much more polymorphic than coding regions and 3' regions. The coding regions had the lowest polymorphism (Fig 3.1). The degeneration of *Fad2* in *D. simulans* might begin from promoter region and resulting nonfunctionalization.

Comparing with *D. melanogaster* and *D. sechellia*, the 5' flanking regions in *D. simulans* (1141 bp) were longer. In *D. simulans*, the 5' flanking region had many indel fragments (less than 15 bp). The 3' flanking region showed large deletions including one 23 bp and one 60 bp deletions. The large-deletion locates in the same region with A-rich sequence in *D. melanogaster* and *D. sechellia*. However, I do not know the effects of those deletions on the *Fad2* transcription.

Tajima's *D*, Fu and Li's, and Fay and Wu's tests were used to test the neutrality of *Fad2* in *D. simulans* populations (Table 3.1A). All values except Fay and Wu's results

of total length and 5’ regions were negative but not significant different from zero and did not reject the null hypothsis of neutrality for *Fad2* plus flanking regions. The positive results of Fay and Wu’s tests in total and 5’ flanking regions might be caused by the excessive low frequency. I also separated the strains into Africa and non-Africa groups to perform the tests (Table 3.1B-C) though the sample sizes were not equal (13: 6). All values except 5’ and 3’ flanking regions were negative but not significant. The positive results of 5’ flanking regions in non-Africa strains were obtained and different from Africa strains (all negative values). Only the 3’ flanking regions in non-Africa strains gained significant negative results in Fu and Li’s tests. The site frequencies of coding regions in all strains, coding and 3’ flanking regions in Africa strains might be too low or too high to perform Fay and Wu’s test.

Table 3.1. The Tajima's *D*, Fu and Li's and Fay and Wu's tests results

A: All strains (N= 20)

	S	π	Tajima's *D*	Fu and Li's *D*	Fu and Li's *F*	Fay and Wu
Total	162	0.0132	-1.3208^{ns}	-1.4656^{ns}	-1.7204^{ns}	0.2269^{ns}
5'region	113	0.0239	-1.1391^{ns}	-1.2139^{ns}	-1.4299^{ns}	0.1848^{ns}
3'region	20	0.0083	-1.6642^{ns}	-1.1678^{ns}	-1.5408^{ns}	-0.0851^{ns}
Coding	29	0.0047	-1.6020^{ns}	-1.9495^{ns}	-2.2367^{ns}	-

B: Africa strains (N= 13)

	S	π	Tajima's *D*	Fu and Li's *D*	Fu and Li's *F*	Fay and Wu
Total	142	0.0135	-1.3597^{ns}	-1.8611^{ns}	-2.0880^{ns}	-0.2919^{ns}
5'region	99	0.0238	-1.2480^{ns}	-1.7148^{ns}	-1.9137^{ns}	0.1210^{ns}
3'region	16	0.0086	-1.4642^{ns}	-2.0722^{ns}	-2.2492^{ns}	-
Coding	27	0.0055	-1.5448^{ns}	-1.6797^{ns}	-2.0007^{ns}	-

C: Non-Africa strains (N= 6)

	S	π	Tajima's *D*	Fu and Li's *D*	Fu and Li's *F*	Fay and Wu
Total	75	0.0129	-0.1842^{ns}	-0.2372^{ns}	-0.2716^{ns}	0.2321^{ns}
5'region	57	0.0246	0.08391^{ns}	0.0730^{ns}	0.1032^{ns}	0.4689^{ns}
3'region	11	0.0096	-1.4448^{ns}	-1.9818*	-2.2263*	0.4276^{ns}
Coding	7	0.0028	-0.2513^{ns}	0.5866^{ns}	0.4739^{ns}	-1.2776^{ns}

GD14271 was a mosaic sequence and only used in all strains analysis and W75 was excluded in all tests. "-" indicates the site frequency is too small or too high to analyze.

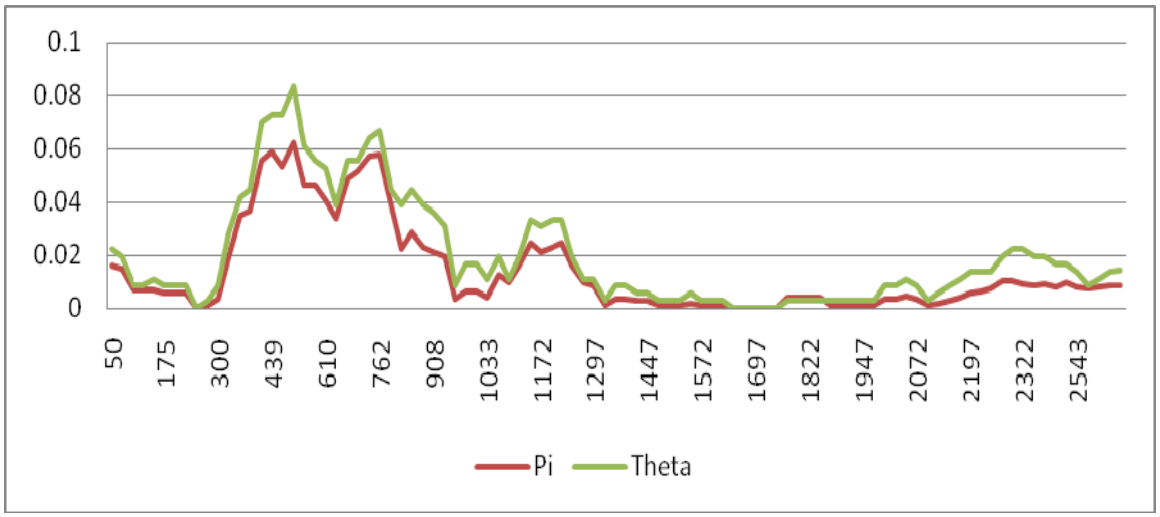


Fig 3.1. The nucleotide diversity *Fad2* with sliding window size of 2690 bp in *D. simulans* 1-1141 bp: 5' flanking region; 1042-2209 bp: coding region; 2210-2690 bp: 3' flanking region.

Most of the neutrality test results were negative but not significant. There were probably a large amount of rare or recent mutations accumulating. However, the excess of rare mutations mean positive selection, and recent mutations were likely to be caused by positive selection, background selection and population demographic factors (Charlesworth, Morgan, and Charlesworth 1993). I further conducted the frequency spectrum of *Fad2* in *D. simulans* populations (Fig 3.2A). The frequency spectrum revealed that there were a large amount of singleton mutations accumulating in *Fad2* sequence of *D. simulans* (including 5' and 3' flanking regions). There was also some excessive mid-frequency (8) and high-frequency allels (19 and 20). The frequency spectrum of Africa strains showed similar patterns: excessive mid-frequency (9) and high-frequency allels (12) (Fig 3.2B). The all site frequencies of non-Africa strains were higher than expected (Fig 3.2C).

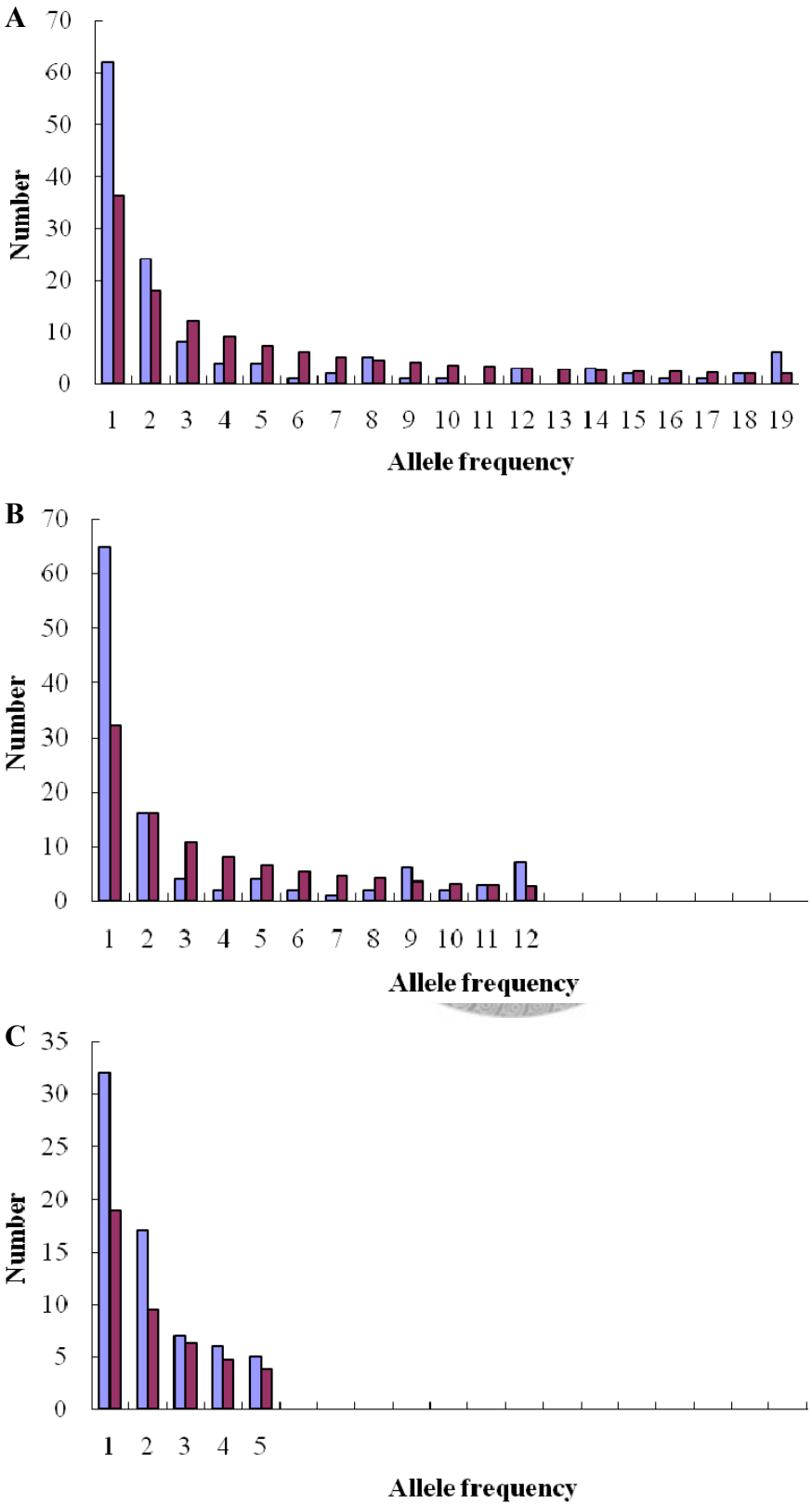


Fig 3.2. Frequency spectrum of *Fad2* alleles in *D. simulans* strains. (A): All strains, N=20, (B): Africa stains, N= 13, (C): Non-Africa strains, N=6. The expected spectrum is estimated according to Watterson's (1975) method. (W75 was excluded)

The McDonald and Kreitman test result showed that *Fad2* coding regions in *D. simulans* did not reject the neutral theory (G= 1.7050, *P*= 0.1916) (Table 3.2). The ratio of silent to replacement sites was close (14:16) within species.

Table 3.2. The McDonald and Kreitman test result for *Fad2* in *D. simulans* (*Fad2* in *D. melanogaster* as outgroup)

	Fixed	Polymorphic	
Silent	32	14	*G*= 1.7050
Replacement	20	16	*P*= 0.1916

Fad2* in *D. yakuba

After surveying 20 *D. yakuba* strains (including *GE21776* in TSC# 14021-0261.01), I found a large number of indels accumulated on *Fad2* (Fig 3.3). The premature stop codon was shown in two strains (CY01 and CY03), one bp insertion in three strains (CY02, CY04 and CY07), two bp insertion in one strain (CY26), five bp insertion in one strain (CY20), 14 bp deletion in one strain (CT01), 12 bp deletion in four strains (CY04, CY09, CY17 and CY23), 664 bp deletion in two strains (CY09 and CY22), four bp deletion in one strain (CY12), 101 bp deletion in two strains (CY13 and CY 23), 84 bp deletion in one strain (CY13), 249 bp deletion in one strain (CY18), 98 bp deletion in two strains (CY21 and CY24), 38 bp deletion in one strain (CY26) and three strains (CY04, CY07 and CY23) had no stop codon. The *D. yakuba* genomic project released *Fad2* sequence (*GE21776* from strain TSC# 14021-0261.01) possessed a 10 bp deletion and one annotated 32 bp intron. The annotated intron sequences were highly conserved among strains. Only three strains (CY08, CY16 and CY19) have complete *Fad2* sequences. Moreover, *Fad2* was not expressed in CY19 (Fig 2.7). Taking the sequence variation and no expression into consideration, *Fad2* in *D. yakuba* might have become a pseudogene. The 5' flanking regions of *Fad2* in *D. yakuba* also showed great divergence with *D. melanogaster*, *D. simulans*, and *D. sechellia* ones. The flanking regions had a lot of deletions which mainly locate within the 500-bp before the

start codon of *Fad2*. The 5’ and 3’ flanking regions were conserved among strains except CY22 which possessed one 800 bp insertion at 5’ flanking region.

Fad2* in *D. teissieri

Four out of sixteen strains, including TUZ11, TUZ35, TUZ125 and T128.2 possessed premature stop codon at the same position (Fig 3.3). TUZ35, TUZ125 and T128.2 also had a three bp deletion and T128.2 had an additional two bp deletion. TUZ11, TUZ35 and TUZ125 were from Tanzania, and T128.2 was received from Dr. Manyuan Long. There was no degeneration in other 12 strains. The result of The McDonald and Kreitman test in *D. teissieri* (exclude degenerated ones) was significant (G= 11.4060, *P*= 0.0007) (Table 3.4A). The interspecific result showed that the replacement mutation number was close to silent one (56:60). Tajima's D and Fu and Li's and Fay and Wu's tests showed negative but not significant results for coding regions (Table 3.3B).

Table 3.3. The McDonald and Kreitman test result (A) and the Tajima's *D* and Fu and Li's tests results (B) for *Fad2* in *D. teissieri*. (*GG15465* in *D. erecta* as outgroup)

(A)

	Fixed	Polymorphic	
Silent	56	63	*G*= 11.4060
Replacement	60	25	*P*= 0.0007

(B)

	S	π	Tajima's *D*	Fu and Li's *D*	Fu and Li's *F*	Fay and Wu's H
Coding	86	0.02584	-0.3558^{ns}	-0.3400^{ns}	-0.3820^{ns}	-0.3986^{ns}

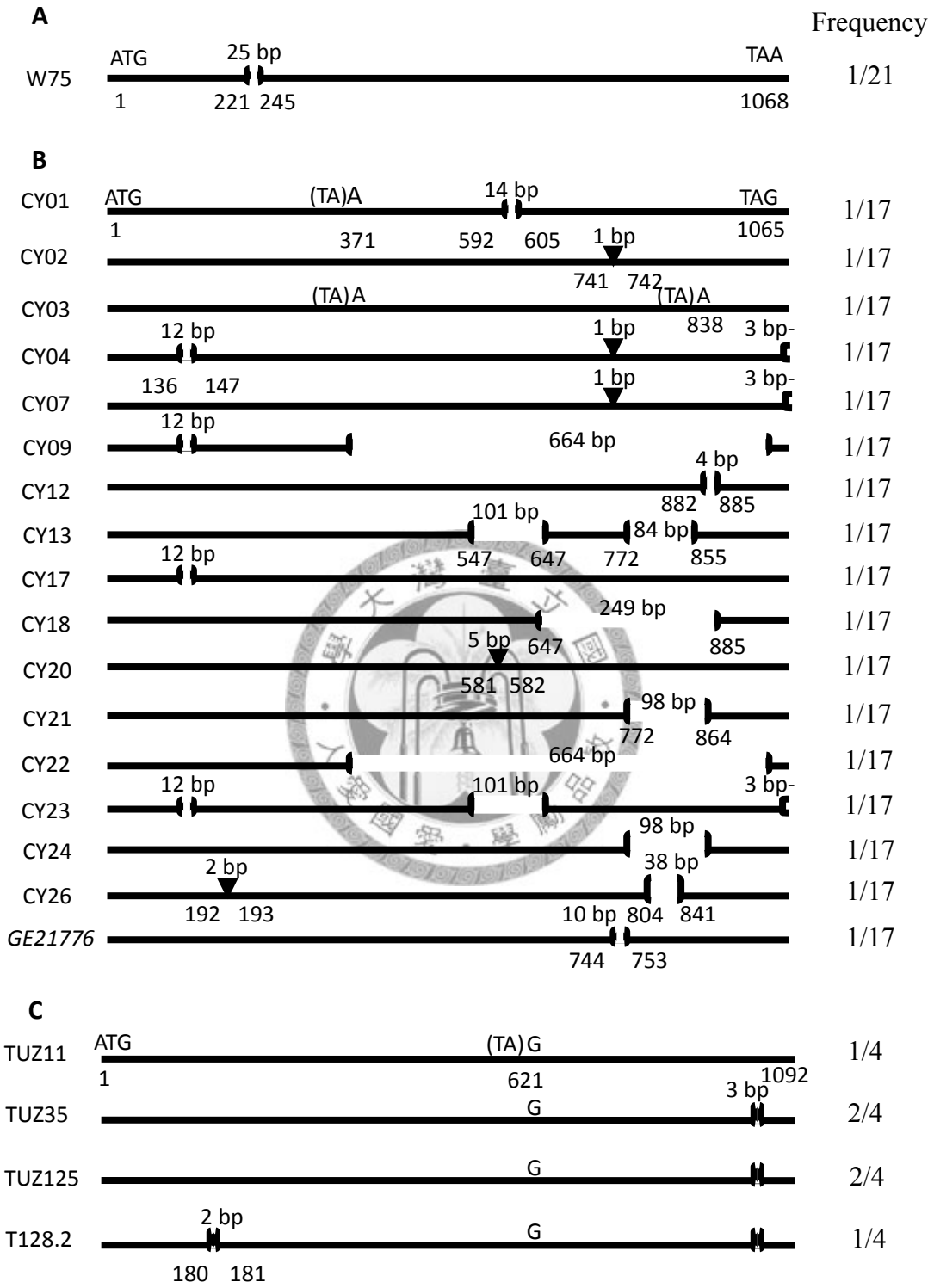


Fig 3.3. The pseudogenization of *Fad2* in (A) *D. simulans*, (B) *D. yakuba* and (C) *D. teissieri*. Each line represents one *Fad2* allele. CY means "*D. yakuba* from Cameroon".

3.4.

Discussion

As only one out of 21 *Fad2* alleles in *D. simulans* became degenerated, the gene might be silenced recently as suggested in Lengendre et al. (2008). According to previous study (Legendre et al. 2008), the divergence and additional sequences with inhibitory elements in the 5'-untranslated region might cause the silencing of *Fad2* in *D. simulans*. Since *Fad2* is related with sex-specific expression and sex pheromone production, the evolution rate might be fast (Proschel, Zhang, and Parsch 2006). The results of neutrality test were not statistically significant which means *Fad2* probably was not a target of selection in *D. simulans*. The ratio of synonymous to nonsynonymous mutations is close to one; it can be explained as relaxation of selection, which means *Fad2* in *D. simulans* might be under selection before. The same phenomenon is also observed in human bitter taste receptor (TAS2R) genes (Wang, Thomas, and Zhang 2004). *AtHVA22d and AtHVA22e* which correlated with plant pheromone abscisic acid and environment stress in *Arabidopsis thaliana* also experienced such selection relaxation. (Chen et al. 2004). However, in *D. simulans*, it would be preferred to degenerate *Fad2* because the production of diene can suppress the courtship behavior from *D. simulans* males (Legendre et al. 2008). The degeneration of 5' flanking regions of *Fad2* in *D. simulans* is much than the coding regions and they

also possess highly polymorphism. The neutrality test results were different between the Africa and non-Africa strains, the 3' flanking regions in non-Africa strains were statistically significant suggesting they possessed more mid-frequency alleles. However, the distance between *Fad2* and 3' flanking gene (*CG32079*) is very short (less than 1000 bp). The statistically significant might represent the 5' flanking region of *CG32079* not the 3' flanking region of *Fad2*. The test results of 5' flanking regions in non-Africa strains (all positive values) are also different from Africa ones (all negative values) that indicates the 5' flanking regions in non-Africa strains possess less mutations or population size changed (decrease or bottleneck). The *Fad2* frequency spectrum of *D. simulans* shows there is also some excessive mid-frequency (8) and high-frequency alleles (19 and 20) that the observed values are higher than expected ones (8: 7.5, 19: 3.2 and 20: 4.2) thought they were not significant. The excessive mid-frequency might indicate the difference between Africa and non-Africa strains because most of them are occurred in Africa strains because most of them were occurred in Africa strains.

Comparing to *D. simulans*, *D. yakuba* showed a higher level of degeneration in *Fad2*. *D. yakuba* showed 16 indels in 17 strains whereas *D. simulans* only one frameshift mutation in one strain. This different degeneration level is probably due to different evolution time, *D. yakuba* has existed at least six million years since the

divergence with *D. teissieri* (Hey and Kliman 1993; Lachaise and Silvain 2004; Tamura, Subramanian, and Kumar 2004) whereas *D. simulans* was only 2.5 millison years. The analysis of 5' flanking regions in *D. simulans* and *D. yakuba* showed the degeneration was not restricted in coding regions. The *Fad2* 5' flanking regions of *Fad2* in both species possessed many indels comparing with functional ones. The 5' flanking regions possessed transcription factor binding sites; indels can disrupt those sites and result in nonfunctionalization.

In *D. teissieri* populations, they showed only two nonsense mutations in three strains from Tanzania and one strain from Dr. Manyuan Long. The divergence time between *D. teissieri* and *D. yakuba* is about six million years ago. However, *Fad2* in *D. yakuba* possessed more serious degeneration than *D. teissieri* ones. *D. yakuba* genome carries numerous chromosome rearrangements and the study of chromosome inversion shows that it evolved substantially faster at the chromosomal level than *D. melanogaster* (28:1) (Lemeunier et al. 1986; Ranz et al. 2007). The accumulation of a large amount of deleterious mutations probably is caused by the higher evolution rate and chromosome rearrangements in a short time or pseudogenized for a long time. I do not know whether the courtship in *D. teissieri* and *D. yakuba* will be effect by diene products like *D. simulans*. If dienes paly an anti-aphrodisiatics role in them, the degeneration of *Fad2* will be favored and results faster evolution rate. *Fad2* in *D.*

teissieri showed similar degeneration pattern to *D. simulans*. The McDonald and Kreitman test result of *D. teissieri* also show *Fad2* in *D. teissieri* possessed excessive polymorphism among strains and faster evolution rate. The pseudogeniztion of *Fad2* in *D. teissieri* probably also occurred recently. However, I did not analize the 5' flanking region of *Fad2* in *D. teissieri*, the degeneration might be happened to there much earlier than the coding region.

A similar finding was *desat2* in cosmopolitan strains of *D. melanogaster*, it is a non-expressed gene but does not accumulated very few deleterious mutations (Greenberg et al. 2006). In European and Asian corn borers (*Ostrinia nubilalis* and *Ostrinia furnacalis*, respectively), there was only Δ11 and Δ14 desaturation involved in the pheromone production (Zhao, Lofstedt, and Wang 1990; Ishikawa, Takanashi, and Huang 1999; Ishikawa et al. 1999; Ma and Roelofs 2002). These two species diverged one million years ago (Roelofs and Rooney 2003). Both of them possessed several duplicated Δ11 and Δ14 *desaturase* genes that were generated by a fusion event with retroposon. Those cryptic *desaturase* genes showed a high identity with the published Δ11 and Δ14 *desaturases* in European and Asian corn borers. However, only some of them were intact and others were neither with incomplete number of exons and intron or truncated or multiple nonsense mutations. Their studies showed that those genes with intact reading frames probably still maintained their functional integrity and existed in

the genome for a relatively long evolutionary time. They also found that those genes showed significant higher d_S value than d_N and was similar with the pattern of purifying selection. Those duplicated *desaturase* genes potentially served as raw material from which new pheromone blends could arise if the genes were co-opted into sex pheromone biosynthesis pathways (Xue et al. 2007).

The evolution of *Fad2* in *D. teissieri* is different from *D. simulans* and *D. yakuba*. The common ancestor of *D. teissieri* and *D. yakuba* probably possesses functional *Fad2* and is likely to maintain it for a long time in *D. teissieri*. However, the role of functional *Fad2* in *D. teissieri*, a sexually monomorphic CHs *Drosophila* species within *D. melanogaster* subgroup, is still unknown. It might serve as a potential speciation factor or a element for environment resistance because the CHs components can affect the environment adaptation (Ferveur 2005). The intra-specific differentiation between two geographically isolated *D. teissieri* populations, Congo and Zimbabwe strains had been shown at songs, sexual activity and wings (Joly, Bizat, and Paillette 1997). I did not analyze *Fad2* in Zimbabwe strains but all Congo stains I used (TBra8, TBra11, TBra12, TBRZ8 and TBRZ17) possessed intact coding sequences. Tanzania is propinquity to Zimbabwe and both of them are geologically distinct from Congo. It is interesting to see if *Fad2* in *D. teissieri* Zimbabwe strains share similar degeneration patterns with Tanzania ones.

Pseudogenization can result the change of female pheromone production, accompanied by a shift in male detection system as exemplified by the case of the European corn borers and Asian corn borers (Roelofs and Rooney 2003). 7,11-HD served as aphrodisiatics in *D. melanogaster* and anti-aphrodisiatics in *D. simulans* (Legendre et al. 2008). Pseudogenization of *Fad2* probably is involved in the speciation between them because dienes, the production of *Fad2,* can result the pre-mating isolation. The effect of dienes in courtship behavior of *D. yakuba* and *D. teissieri* is still unknown.

I found degenerated *Fad2* sequences among three sexually monomorphic CHs *D. melanogaster* subgroup species. However, the degrees of degeneration are variable among them especially between *D. teissieri* and *D. yakuba*. Their pseudogenization time might be different which means the evolution of *Fad2* happened independently among *D. simulans*, *D. teissieri* and *D. yakuba* though they might also be affected by different evolution rates among species. The lineage specific evolution of *Fad2* occurred in those closely related species. This phenomenon can reflect the distinct evolution of pheromone among species within *D. melanogaster* subgroup.

Conclusion

Fad2 is a *Drosophila*-specific Δ9 *desaturase* gene and it is created by retrotransposition (Bai et al. 2007). Its evolution may be variable among *Drosophila* speices that they possess variable expression patterns and speices-specific paralogs. The distinct evolution among species may reflect the species-specific sex pheromone production.

Fad2 plays an important role in the production of female-specific sex pheromone in sexually dimorphic CHs species within *D .melanogaster* subgroup. In this subgroup, sexually monomorphic CHs species possess non-expressed and pseudogenized *Fad2.* The nonfunctionalization of *Fad2* may be an advantage because the female-specific sex pheromone may be anti-aphrodisiatics to sexually monomorphic CHs species.

The large survey of the expression patterns among strains form each species shall be performed to find out the strain variation. The expression of *Fad2* in different *Drosophila* developmental stages should also be investigated. The complete degeneration pattern might be elucidated by investigating other species within *D. melanogaster* subgroup with sexually monomorphic CHs including *D. mauritiana*, *D. orena* and *D. santomea.* Those studies will give a completely view of molecular evolution of *desaturase* gene duplication in *D. melanogaster* species subgroup.


國立臺灣大學
人愛國愛學勵品教

References

An W, and Wensinck PC. 1995. Three protein binding sites form an enhancer that regulates sex- and fat body-specific transcription of Drosophila yolk protein genes. EMBO J. **14**:1221-1230.

Antony C, Davis T, Davis TL, Pechine JM, and Jallon JM. 1985. Compared behavioral responses of male *Drosophila melanogaster* (Canton S) to natural and synthetic aphrodisiacs. J Chem Ecol. **11**:1617-1629.

Antony C, and Jallon JM. 1982. The chemical basis for sex recognition in *Drosophila melanogaster*. J Insect Physiol. **28**:873-880.

Bai Y, Casola C, Feschotte C, and Betran E. 2007. Comparative genomics reveals a constant rate of origination and convergent acquisition of functional retrogenes in *Drosophila*. Genome Biol. **8**:R11.

Baker BS, Burtis K, Goralski T, Mattox W, and Nagoshi R. 1989. Molecular genetic aspects of sex determination in *Drosophila melanogaster*. Genome. **31**:638-645.

Bjostad LB, Wolf WA, and Roelofs WL. 1987. Pheromone biosynthesis in *Lepidopterans*: desaturalion and chain-shortening *in* Prestwich.GD and Blomquist.GJ eds. Pheromone Biochemistry. Arcademic Press, New York.

Braverman JM, Hudson RR, Kaplan NL, Langley CH, and Stephan W. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics. **140**:783-796.

Burland TG. 2000. DNASTAR's Lasergene sequence analysis software. Methods Mol Biol. **132**:71-91.

Burtis KC, Thummel CS, Jones CW, Karim FD, and Hogness DS. 1990. The *Drosophila* 74EF early puff contains E74, a complex ecdysone-inducible gene that encodes two ets-related proteins. Cell. **61**:85-99.

Charlesworth B, Morgan MT, and Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics. **134**:1289-1303.

Chen CN, Chiang YC, Ho TH, Schaal BA, and Chiang TY. 2004. Coalescent processes and relaxation of selective constraints leading to contrasting genetic diversity at paralogs *AtHVA22d* and *AtHVA22e* in *Arabidopsis thaliana*. Mol Phylogenet Evol. **32**:616-626.

Chertemps T, Duportets L, Labeur C, Ueda R, Takahashi K, Saigo K, and Wicker-Thomas C. 2007. A female-biased expressed elongase involved in long-chain hydrocarbon biosynthesis and courtship behavior in *Drosophila melanogaster*. Proc Natl Acad Sci USA. **104**:4273-4278.

Chertemps T, Duportets L, Labeur C, Ueyama M, and Wicker-Thomas C. 2006. A female-specific desaturase gene responsible for diene hydrocarbon biosynthesis and courtship behaviour in *Drosophila melanogaster*. Insect Mol Biol.

15:465-473.

Cho HP, Nakamura M, and Clarke SD. 1999a. Cloning, expression, and fatty acid regulation of the human delta-5 desaturase. J Biol Chem. **274**:37335-37339.

Cho HP, Nakamura MT, and Clarke SD. 1999b. Cloning, expression, and nutritional regulation of the mammalian Delta-6 desaturase. J Biol Chem. **274**:471-477.

Choi T, Cho N, Oh Y, Yoo M, Matsukage A, Ryu Y, Han K, Yoon J, and Baek K. 2000. The DNA replication-related element (DRE)-DRE-binding factor (DREF) system may be involved in the expression of the *Drosophila melanogaster* TBP gene. FEBS Lett. **483**:71-77.

Clewley JP. 1997. GENEMAN of LASERGENE. Methods Mol. Biol. **70**:189-196.

Clewley JP. 1995. Macintosh sequence analysis software. DNAStar's LaserGene. Mol Biotechnol. **3**:221-224.

Clewley JP, and Arnold C. 1997. MEGALIGN. The multiple alignment module of LASERGENE. Methods Mol Biol. **70**:119-129.

Cline TW. 1993. The *Drosophila* sex determination signal: how do flies count to two? Trends Genet. **9**:385-390.

Cobb M, and Jallon JM. 1990. Pheromones, mate recognition and courtship stimulation in the *Drosophila melanogaster* species sub-group. Anim Behav. **39**:1058-1069.

Coschigano KT, and Wensink PC. 1993. Sex-specific transcriptional regulation by the male and female doublesex proteins of *Drosophila*. Genes Dev. **7**:42-54.

Coyne JA. 1996. Genetics of differences in pheromonal hydrocarbons between *Drosophila melanogaster* and *D. simulans*. Genetics. **143**:353-364.

Coyne JA, and Charlesworth B. 1997. Genetics of a pheromonal difference affecting sexual isolation between *Drosophila mauritiana* and *D. sechellia*. Genetics. **145**:1015-1030.

Coyne JA, and Orr HA. 2004. Speciation, Sunderland, MA.

Dallerac R, Labeur C, Jallon JM, Knipple DC, Roelofs WL, and Wicker-Thomas C. 2000. A delta 9 desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in *Drosophila melanogaster*. Proc Natl Acad Sci USA. **97**:9449-9454.

Dugdale JS. 1997. Insect Pheromone Research. Chapman & Hall, New York.

Edney EB. 1967. Water balance in desert arthropods. Despite their small size, arthropods may be highly adapted for life in xeric conditions. Science. **156**:1059-1066.

England BP, Heberlein U, and Tjian R. 1990. Purified *Drosophila* transcription factor, Adh distal factor-1 (Adf-1), binds to sites in several *Drosophila* promoters and activates transcription. J Biol Chem. **265**:5086-5094.

Fang S, Takahashi A, and Wu CI. 2002. A mutation in the promoter of *desaturase 2* is

correlated with sexual isolation between *Drosophila* behavioral races. Genetics. **162**:781-784.

Fay JC and Wu CI. Hitchhiking under positive Darwinian selection. Genetics. **155**: 1405-1413.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. **39**:783-791.

Ferveur JF. 2005. Cuticular hydrocarbons: their evolution and roles in *Drosophila* pheromonal communication. Behav Genet. **35**:279-295.

Ferveur JF, Cobb M, Boukella H, and Jallon JM. 1996. World-wide variation in *Drosophila melanogaster* sex pheromone: behavioural effects, genetic bases and potential evolutionary consequences. Genetica. **97**:73-80.

Ferveur JF, and Sureau G. 1996. Simultaneous influence on male courtship of stimulatory and inhibitory pheromones produced by live sex-mosaic *Drosophila melanogaster*. Proc Biol Sci. **263**:967-973.

Fitch WM. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. **20**:406–416.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, and Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. **151**:1531-1545.

Fu YX. 1996. New statistical tests of neutrality for DNA samples from a population. Genetics. **143**:557-570.

Fu YX, and Li WH. 1993. Statistical tests of neutrality of mutations. Genetics. **133**:693-709.

Ginder GD, Gnanapragasam MN, and Mian OY. 2008. The role of the epigenetic signal, DNA methylation, in gene regulation during erythroid development. Curr Top Dev Biol. **82**:85-116.

Gleason JM, Jallon JM, Rouault JD, and Ritchie MG. 2005. Quantitative trait loci for cuticular hydrocarbons associated with sexual isolation between *Drosophila simulans* and *D. sechellia*. Genetics. **171**:1789-1798.

Greenberg AJ, Moran JR, Fang S, and Wu CI. 2006. Adaptive loss of an old duplicated gene during incipient speciation. Mol Biol Evol. **23**:401-410.

Greenspan RJ, and Ferveur JF. 2000. Courtship in *Drosophila*. Annu Rev Genet. **34**:205-232.

Harrison PM, Echols N, and Gerstein MB. 2001. Digging for dead genes: an analysis of the characteristics of the pseudogene population in the *Caenorhabditis elegans* genome. Nucleic Acids Res. **29**:818-830.

Harrison PM, Hegyi H, Balasubramanian S, Luscombe NM, Bertone P, Echols N, Johnson T, and Gerstein M. 2002. Molecular fossils in the human genome:

identification and analysis of the pseudogenes in chromosomes 21 and 22. Genome Res. **12**:272-280.

Higashi S, and Murata N. 1993. An in vivo study of substrate specificities of acyl-lipid desaturases and acyltransferases in lipid synthesis in *synechocystis* PCC6803. Plant Physiol. **102**:1275-1278.

Hollocher H, Ting CT, Wu ML, and Wu CI. 1997. Incipient speciation by sexual isolation in *Drosophila melanogaster*: extensive genetic divergence without reinforcement. Genetics. **147**:1191-1201.

Hsu T, Gogos JA, Kirsh SA, and Kafatos FC. 1992. Multiple zinc finger forms resulting from developmentally regulated alternative splicing of a transcription factor gene. Science. **257**:1946-1950.

Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. Proc Biol Sci. **256**:119-124.

Hughes AL, and Nei M. 1989. Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. Mol Biol Evol. **6**:559-579.

Ishikawa Y, Takanashi T, and Huang YP. 1999. Comparative studies on the sex pheromones of *Ostrinia spp*. in Japan: the burdock borer, *Ostrinia zealis*. Chemoecology. **9**:25-32.

Ishikawa Y, Takanashi T, Kim C, Hoshizaki S, Tatsuki S, and Huang YP. 1999. *Ostrinia spp*. in Japan: their host plants and sex pheromones. Entomol Exp Appl. **91**:237-244.

Jallon JM. 1984. A few chemical words exchanged by *Drosophila* during courtship and mating. Behav Genet. **14**:441-478.

Jallon JM, and David JR. 1987. Variations in cuticular hydrocarbons among the eight soecies of the *Drosophila melanogaster* subgroup. Evolution. **41**:294-302.

Jallon JM, and Hotta Y. 1979. Genetic and behavioral studies of female sex appeal in *Drosophila*. Behav Genet. **9**:257-275.

Jallon JM, and Pechiné JM. 1989. Une autre race chimique de *Drosophila melanogaster* en Afrique. CR Acad Sci Paris. **309**:1551-1556.

Jallon JM, and Wicker-Thomas C. 2003. Genetic study on pheromone production in *Drosophila in* G. Blomquist, and R. Vogt, eds. Insect Pheromone Biochemistry and Molecular Biology Elsevier Academic Press, London.

Jensen RA. 1976. Enzyme recruitment in evolution of new function. Annu Rev Microbiol. **30**:409-425.

Joly D, Bizat N, and Paillette M. 1997. Differentiation of dialects and courtship strategies in allopatric populations of *Drosophila teissieri*. J. Insect Physiol. **43**:809-814.

Jurenka RA. 2003. Biochemistry of female moth sex pheromones *in* Blomquist GJ and RG eds. Insect Pheromone Biochemistry and Molecular Biology. Elsevier Academic Press, London.

Kaplan NL, Hudson RR, and Langley CH. 1989. The "hitchhiking effect" revisited. Genetics. **123**:887-899.

Klages N, and Strubin M. 1995. Stimulation of RNA polymerase II transcription initiation by recruitment of TBP in vivo. Nature. **374**:822-823.

Knipple DC, Rosenfield CL, Nielsen R, You KM, and Jeong SE. 2002. Evolution of the integral membrane desaturase gene family in moths and flies. Genetics. **162**:1737-1752.

Kohany O, Gentles AJ, Hankus L, and Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics. **7**:474.

Kuhn EJ, and Geyer PK. 2003. Genomic insulators: connecting properties to mechanism. Curr Opin Cell Biol. **15**:259-265.

Labeur C, Dallerac R, and Wicker-Thomas C. 2002. Involvement of desat1 gene in the control of *Drosophila melanogaster* pheromone biosynthesis. Genetica. **114**:269-274.

Laney JD, and Biggin MD. 1992. *zeste*, a nonessential gene, potently activates Ultrabithorax transcription in the *Drosophila* embryo. Genes Dev. **6**:1531-1541.

Lee KN, Pariza MW, and Ntambi JM. 1996. Differential expression of hepatic stearoyl-CoA desaturase gene 1 in male and female mice. Biochim Biophys Acta. **1304**:85-88.

Legendre A, Miao XX, Da Lage JL, and Wicker-Thomas C. 2008. Evolution of a desaturase involved in female pheromonal cuticular hydrocarbon biosynthesis and courtship behavior in *Drosophila*. Insect Biochem Mol Biol. **38**:244-255.

Lemeunier F, JR David, Tsacas L, and Ashburner M. 1986. The *melanogaster* species group. Pp. 5 *in* M. Ashburner, H. L. Carson, and J. N. Thompson, eds. The Genetics and Biology of *Drosophila*. Academic Press, New York.

Leptin M. 1991. twist and snail as positive and negative regulators during *Drosophila* mesoderm development. Genes Dev. **5**:1568-1576.

Long M. 2001. Evolution of novel genes. Curr Opin Genet Dev. **11**:673-680.

Ma PW, and Roelofs WL. 2002. Sex pheromone gland of the female European corn borer moth, *Ostrinia nubilalis* (Lepidoptera, Pyralidae): ultrastructural and biochemical evidences. Zoolog Sci. **19**:501-511.

Mas F, and Jallon JM. 2005. Sexual isolation and cuticular hydrocarbon differences between *Drosophila santomea* and *Drosophila yakuba*. J Chem Ecol. **31**:2747-2752.

Maynard Smith J, and Haigh J. 1974. The hitch-hiking effect of a favorable gene. Genet Res. **23**:23-35.

Mayr E. 1963. Animal Species and Evolution. Belknap Press, Cambridge, MA.

McDonald JH, and Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature. **351**:652-654.

Miller GT, and Pitnick S. 2002. Sperm-female coevolution in *Drosophila*. Science. **298**:1230-1233.

Moto K, Suzuki MG, Hull JJ, Kurata R, Takahashi S, Yamamoto M, Okano K, Imai K, Ando T, and Matsumoto S. 2004. Involvement of a bifunctional fatty-acyl desaturase in the biosynthesis of the silkmoth, *Bombyx mori*, sex pheromone. Proc Natl Acad Sci USA. **101**:8631-8636.

Nei M, Gu X, and Sitnikova T. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci USA. **94**:7799-7806.

Nei M, and Hughes AL. 1992. Proceedings of the 11th Histocompatibility Workshop and Conference. Oxford Univ. Press, Oxford.

Neilson R. 2005. Molecular Signatures of Natural Selection. Annu Rev Genet. **39**:197-218.

Nowak MA, Boerlijst MC, Cooke J, and Smith JM. 1997. Evolution of genetic redundancy. Nature. **388**:167-171.

Nuesslein-Volhard C, Wieschaus E, and Kluding H. 1984. Mutations affecting the pattern of larval cuticle in *Drosphila melanogaster*. I. Zygotic loci on the second chromosome. Roux's Arch Dev Biol. **193**:267-282.

Orgel LE. 1977. Gene-duplication and the origin of proteins with novel functions. J Theor Biol. **67**:773.

Parker CS, and Topol J. 1984. A *Drosophila* RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. Cell. **36**:357-369.

Patterson HEH. 1985. The recognition of species *in* E. S. Vrba, ed. Species and Speciation. Monogr., Transvaal Mus.

Pechine JM, Antony C, and Jallon JM. 1988. Precise characterization of cuticular compounds in young *Drosophila* by mass spectrometry. J Chem Ecol. **14**:1071-1085.

Pennanec' h M, Ferveur JF, Pho DB, and Jallon JM. 1991. Insect fatty acid related pheromones: a review of their biosynthesis, hormonal regulation and genetic control. Ann Soc Entomol Fr. **27**:245-263.

Pennanec' h M, Bricard L, Kunesh G, and Jallon JM. 1997. Incorporation of fatty acids into cuticular hydrocarbons of male and female *Drosophila melanogaster.* J Insect Physiol. **31**:289-303.

Perez DE, and Wu CI. 1995. Further characterization of the Odysseus locus of hybrid sterility in *Drosophila*: one gene is not enough. Genetics **140**:201-206.

Perkins KK, Dailey GM, and Tjian R. 1988. In vitro analysis of the *Antennapedia* P2 promoter: identification of a new *Drosophila* transcription factor. Genes Dev. **2**:1615-1626.

Pitnick S, Markow T, and Spicer GS. 1999. Evolution of multiple kinds of sperm storage organs in *Drosophila*. Evolution. **53**:1804-1822.

Proschel M, Zhang Z, and Parsch J. 2006. Widespread adaptive evolution of *Drosophila* genes with sex-biased expression. Genetics. **174**:893-900.

Ranz JM, Maurin D, Chan YS, von Grotthuss M, W. Hillier L, Roote J, Ashburner M, and Bergman CM. 2007. Principles of genome evolution in the *Drosophila melanogaster* species group. PLoS Biol. **5**:e152.

Reed JR, Quilici DR, Blomquist GJ, and Reitz RC. 1995. Proposed mechanism for the cytochrome P450-catalyzed conversion of aldehydes to hydrocarbons in the house fly, *Musca domestica*. Biochemistry. **34**:16221-16227.

Reed JR, Vanderwel D, Choi S, Pomonis JG, Reitz RC, and Blomquist GJ. 1994. Unusual mechanism of hydrocarbon formation in the housefly: cytochrome P450 converts aldehyde to the sex pheromone component (Z)-9-tricosene and CO_2. Proc Natl Acad Sci USA. **91**:10000-10004.

Riggleman B, Wieschaus E, and Schedl P. 1989. Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a *Drosophila* segment polarity gene. Genes Dev. **3**:96-113.

Rivera-Pomar R, Lu X, Perrimon N, Taubert H, and Jackle H. 1995. Activation of posterior gap gene expression in the *Drosophila* blastoderm. Nature. **376**:253-256.

Rodriguez F, Hallahan DL, Pickett JA, and Camps F. 1992. Characterization of the Δ[11]-palmitoyl-CoA-desaturase from *Spodoptera littoralis* (Lepidoptera : noctuidae). Insect Biochem Mol Biol. **22**:143-148.

Roelofs WL, and Rooney AP. 2003. Molecular genetics and evolution of pheromone biosynthesis in *Lepidoptera*. Proc Natl Acad Sci. USA. **100**:9179-9184.

Ronquist FR, and Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**:1572–1574.

Rooney AP, Piontkivska H, and Nei M. 2002. Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. Mol Biol Evol. **19**:68-75.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, and Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19**:2496-2497.

Saitou N, and Nei M. 1987. The neighbor-joining method: a new method for

reconstructing phylogenetic trees. Mol Biol Evol. **4**:406-425.

Sawado T, Hirose F, Takahashi Y, Sasaki T, Shinomiya T, Sakaguchi K, Matsukage A, and Yamaguchi M. 1998. The DNA replication-related element (DRE)/DRE-binding factor system is a transcriptional regulator of the *Drosophila E2F* gene. J Biol Chem. **273**:26042-26051.

Schully SD, and Hellberg ME. 2006. Positive selection on nucleotide substitutions and indels in accessory gland proteins of the *Drosophila pseudoobscura* subgroup. J Mol Evol. **62**:793-802.

Serra M, Pina B, Abad JL, Camps F, and Fabrias G. 2007. A multifunctional desaturase involved in the biosynthesis of the processionary moth sex pheromone. Proc Natl Acad Sci USA. **104**:16444-16449.

Serra M, Pina B, Bujons J, Camps F, and Fabrias G. 2006. Biosynthesis of 10,12-dienoic fatty acids by a bifunctional Delta11 desaturase in *Spodoptera littoralis*. Insect Biochem Mol Biol. **36**:634-641.

Sprecher H. 2000. Metabolism of highly unsaturated n-3 and n-6 fatty acids. Biochim Biophys Acta. **1486**:219-231.

Steinmann-Zwicky M, Amrein H, and Nothiger R. 1990. Genetic control of sex determination in *Drosophila*. Adv Genet. **27**:189-237.

Struhl G, Johnston P, and Lawrence PA. 1992. Control of Drosophila body pattern by the hunchback morphogen gradient. Cell. **69**:237-249.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. **123**:585-595.

Takahashi A, Tsaur SC, Coyne JA, and Wu CI. 2001. The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. Proc Natl Acad Sci USA. **98**:3920-3925.

Thisse B, Messal M, and Perrin-Schmitt F. 1987. The twist gene: isolation of a *Drosophila* zygotic gene necessary for the establishment of dorsoventral pattern. Nucleic Acids Res. **15**:3439-3453.

Thomas HE, Stunnenberg HG, and Stewart AF. 1993. Heterodimerization of the *Drosophila* ecdysone receptor with retinoid X receptor and ultraspiracle. Nature. **362**:471-475.

Thummel CS. 1989. The *Drosophila* E74 promoter contains essential sequences downstream from the start site of transcription. Genes Dev. **3**:782-792.

Thummel CS, Burtis KC, and Hogness DS. 1990. Spatial and temporal patterns of E74 transcription during *Drosophila* development. Cell. **61**:101-111.

Ting CT, Tsaur SC, Sun S, Browne WE, Chen YC, Patel NH, and Wu CI. 2004. Gene duplication and speciation in *Drosophila*: evidence from the *Odysseus* locus. Proc Natl Acad Sci USA. **101**:12232-12235.

Tocher DR, Leaver MJ, and Hodgson PA. 1998. Recent advances in the biochemistry and molecular biology of fatty acyl desaturases. Prog Lipid Res. **37**:73-117.

Toolson EC, and Kuper-Simbron R. 1989. Laboratory evolution of epicuticular hydrocarbon composition and cuticular permeability in *Drosophila pseudoobscura*: effect on sexual dimorphism and thermal-acclimation ability. Evolution. **43**:468-473.

Tsaur SC, Ting CT, and Wu CI. 1998. Positive selection driving the evolution of a gene of male reproduction, Acp26Aa, of *Drosophila*: II. Divergence versus polymorphism. Mol Biol Evol. **15**:1040-1046.

Tsaur SC, and Wu CI. 1997. Positive selection and the molecular evolution of a gene of male reproduction, Acp26Aa of *Drosophila*. Mol Biol Evol. **14**:544-549.

Wagstaff BJ, and Begun DJ. 2007. Adaptive evolution of recently duplicated accessory gland protein genes in desert *Drosophila*. Genetics. **177**:1023-1030.

Wang X, Thomas SD, and Zhang J. 2004. Relaxation of selective constraint and loss of function in the evolution of human bitter taste receptor genes. Hum Mol Genet. **13**:2671-2678.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. **7**:256-276.

Wicker-Thomas C. 2007. Pheromonal communication involved in courtship behavior in Diptera. J Insect Physiol. **53**:1089-1100.

Wicker-Thomas C, Henriet C, and Dallerac R. 1997. Partial characterization of a fatty acid desaturase gene in *Drosophila melanogaster*. Insect Biochem Mol Biol. **27**:963-972.

Wicker-Thomas C, and Jallon JM. 2000. Role of *Enhancer of zeste* on the production of *Drosophila melanogaster* pheromonal hydrocarbons. Naturwissenschaften. **87**:76-79.

Wilkins RC, and Lis JT. 1998. GAGA factor binding to DNA via a single trinucleotide sequence element. Nucleic Acids Res. **26**:2672-2678.

Wolf WA, and Roelofs WL. 1986. Properties of the desaturase enzyme used in cabbage looper moth sex pheromone biosynthesis. Arch Insect Biochem Physiol. **3**.

Wong A, Turchin MC, Wolfner MF, and Aquadro CF. 2008. Evidence for positive selection on *Drosophila melanogaster* seminal fluid protease homologs. Mol Biol Evol. **25**:497-506.

Wu CI, Hollocher H, Begun DJ, Aquadro CF, Xu Y, and Wu ML. 1995. Sexual isolation in *Drosophila melanogaster*: a possible case of incipient speciation. Proc Natl Acad Sci USA. **92**:2519-2523.

Xue B, Rooney AP, Kajikawa M, Okada N, and Roelofs WL. 2007. Novel sex pheromone desaturases in the genomes of corn borers generated through gene

duplication and retroposon fusion. Proc Natl Acad Sci USA. **104**:4467-4472.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. **24**:1586-1591.
Zhang J. 2003. Evolution by gene duplication: an update. Trends Ecol Evol. **18**:292-298.
Zhang J, Rosenberg HF, and Nei M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci USA. **95**:3708-3713.
Zhao C, Lofstedt C, and Wang X. 1990. Sex pheromone biosynthesis in the asian corn borer *Ostrinia furnacalis* (II): Biosynthesis of (E)- and (Z)-12-tetradecenyl acetate involves (delta) desaturation. Arch Insect Biochem Physiol. **15**:57-65.

Appendix

Table S1: the primer list used in PCR and RT-PCR. (*: for PCR only.)

Species	Gene	Primer name	sequence
D. melanogaster	*Fad2*	Fad2_-1138F*	5'-TTATGAACCAGCAAATGAG-3'
D. simulans	*GD14271*	Fad2_-387F*	5'-CAAATAAAAATARCCCCGTAA-3'
D. sechellia	*GM25238*	Fad2_-196R*	5'-AGAAGTTCGCATTYATTAGCA-3'
	S9	Fad2_280F	5'-TGGGCTCATCGCACMTTC-3'
D. yakuba	*GE21776*	Fad2_354R*	5'-GCCCAATAGTAGACCGCATCC-3'
	CY19	Fad2_882F*	5'-TCGGAGGATGTGACCACTAAG-3'
D. santomea	*LAGO14082*	Fad2_986R	5'-TATCGCCCCAACCCCAAACAG-3'
D. erecta	*GG15465*	Fad2_1515R*	5'-AGCCACCAMCTCCTCCAATC-3'
		Fad2_1832R*	5'-GAAGGGTGGAGGTATTAG-3'
		ds-5'	5'-ATGCCACGCAATACCAAAGAC-3'
D. ananassae	*GF24026*	GLEANR8783_ana_165F	5'-CTCGCTGTACGGGGTGTG-3'
		GLEANR8783_ana_1035R	5'-TTGTGAGTGATGGTAGTC-3'
	GF16174	Fad2_280F	5'-TGGGCTCATCGCACMTTC-3'
		GLEANR17446_ana_947R	5'-GTTGGACTCTTTTTCTCAC-3'
	GF18504	GLEANR19760_ana_145F	5'-ATCCTCTTCGCAGTTCTTC-3'
		GLEANR19760_ana_1001R	5'-GGTCCTTGTCTCCCCATC-3'
D. pseudoobscura	*GA20691,*	GA20691_pse_36F	5'-GCGGCAACGGTTCTGAAG-3'
and D.persimilis	*GL15669*	GA20691_pse_983R	5'-GTGTCGGCAATGTCCTCCT-3'
	GA27148,	desatF-B_pse_108F	5'-CACTCTGAAGACGACGGA-3'
	GL23117	desatF-B_pse_822R	5'-TGGTGATAGTTGTGGTAG-3'

	GA27452,	desatF-C_pse_197F	5’-TCTGGCTTATGCTCACAA-3’
	GL22317	desatF-C_pse_1048R	5’-GGTAATGGTCGTGCTATC-3’
D. willistoni	*GK17186*	GLEANR17505_wil_1F	5’-ATGCCACCGAATGCCGAA-3’
		GLEANR17505_wil_981R	5’-CCAGTGCGTATAACCCGT-3’
	GK11373	GLEANR1159_wil_184F	5’-CTATCGGGCTATGGTTTGTG-3’
		GLEANR1159_wil_777R	5’-GTATTGGTTCATTGGTGGGA-3’
	Act5c	Act5C_c713+	5’-TTTGACCGACTACCTGATGAAG-3’
		Act5C_c958-	5’-CAAGCCTCCATTCCCAAGA-3’

Table S2: Desaturase gene homologs in 12 *Drosophila* species.

Desaturase gene	Species	Annotation symbol (Flybase ID)	Genomic location
desat1	*D. melanogaster*	*CG5887* (FBgn0086687)	3R:8270531..8272663
	D. simulans	*GD18837* (FBgn0027803)	3R:13116857..13118963
	D. sechellia	*GM24036* (FBgn0178901)	scaffold_0:13674699..13676859
	D. yakuba	*GE26198* (FBgn0067988)	3R:12593935..12595999
	D. erecta	*GG18949* (FBgn0111155)	scaffold_4770:13348445..13350478
	D. ananassae	*GF17961* (FBgn0094979)	scaffold_13340:7652350..7654184
	D. pseudoobscura	*GA19204* (FBtr0284973)	2:5498572..5500153
	D. persimilis	*GL27316* (FBgn0164897)	scaffold_19:1215069..1216653
	D. willistoni	*GK14407* (FBgn0216413)	scf2_1100000004943:14785806..14787424
	D. mojavensis	*GI10485* (FBgn0133249)	scaffold_6540:27937726..27939979
	D. virilis	*GJ23166* (FBgn0210268)	scaffold_13047:8252067..8254552
	D. grimshawi	*GH21229* (FBgn0128691)	scaffold_15074:1355903..1358096
desat2	*D. melanogaster*	*CG5925* (FBgn0043043)	3R:8262143..8263481
	D. simulans	*GD18836* (FBgn0190352)	3R:13126012..13127348
	D. sechellia	*GM24035* (FBgn0178900)	scaffold_0:13683759..13685095
	D. yakuba	*GE26197* (FBgn0243231)	3R:12585391..12586717
	D. erecta		
	D. ananassae	*GF17959* (FBgn0094977)	scaffold_13340:7642863..7644162
	D. pseudoobscura	*GA19234* (FBgn0079231)	2:5488491..5489718
	D. persimilis	*GL27314* (FBgn0164895)	scaffold_19:1204796..1206118

	D. willistoni	*GK14406* (FBgn0216412)	scf2_1100000004943:14779457..14780725
	D. mojavensis	*GI10484* (FBgn0133248)	scaffold_6540:27929644..27930945
	D. virilis	*GJ23165* (FBgn0210267)	scaffold_13047:8261264..8262614
	D. grimshawi	*GH21218* (FBgn0128680)	scaffold_15074:1348964..1350306
Fad2	*D. melanogaster*	*CG7923* (FBgn0029172)	3L:11016639..11017706
	D. simulans	*GD14271*(FBgn0185956)	3L:10413484..10414551
	D. sechellia	*GM25238* (FBgn0180098)	scaffold_0:3239851..3240915
	D. yakuba	*GE21776* (FBgn0239021)	3L:11044102..11045156
	D. erecta	*GG15465* (FBgn0107716)	scaffold_4784:11026003..11027067
	D. ananassae	*GF16174* (FBgn0093196)	scaffold_13337:12461680..12462747
		GF24026 (FBgn0101020)	scaffold_13340:15668648..15669706
		GF18504 (FBgn0095522)	scaffold_13340:21312697..21313764
	D. pseudoobscura	*GA20691* (FBgn0080685)	XR_group6:6011391..6012452
		GA27148 (FBgn0248519)	2:10728335..10729423
		GA27452 (FBgn0248819)	2:23774445..23775530
	D. persimilis	*GL15669* (FBgn0153273)	scaffold_47:509015..510076
		GL23117 (FBgn0160707)	scaffold_0:9749674..9750762
		GL22317 (FBgn0159909)	scaffold_3:6548217..6549302
	D. willistoni	*GK17186* (FBgn0219185)	scf2_1100000004511:2193764..2194855
		GK11373 (FBgn0213384)	scf2_1100000004762:97180..98253
	D. mojavensis		
	D. virilis		
	D. grimshawi		

CG8630	*D. melanogaster*	*CG8630* (FBgn0038130)	3R:9108124..9109675
	D. simulans	*GD18893* (FBgn0190404)	3R:12316156..12317679
	D. sechellia	*GM24094* (FBgn0178958)	scaffold_0:12894947..12896458
	D. yakuba	*GE26257* (FBgn0243288)	3R:13412487..13414053
	D. erecta	*GG19533* (FBgn0111735)	scaffold_4770:12539421..12541069
	D. ananassae	*GF17588* (FBgn0094606)	scaffold_13340:1662762..1664203
	D. pseudoobscura	*GA21221* (FBgn0081209)	2:9994113..9997403
	D. persimilis	*GL24461* (FBgn0162050)	scaffold_0:10476768..10488896
	D. willistoni	*GK13932* (FBgn0215938)	scf2_1100000004943:6191658..6193093
	D. mojavensis	*GI24327* (FBgn0147050)	scaffold_6540:10145793..10147297
	D. virilis	*GJ10413* (FBgn0197694)	scaffold_12855:3052346..3053856
	D. grimshawi	*GH18426* (FBgn0125894)	scaffold_14906:5753675..5755127
CG9743	*D. melanogaster*	*CG9743* (FBgn0039756)	3R:26022359..26024615
	D. simulans	*GD17212* (FBgn0188775)	3R:25690152..25692391
	D. sechellia	*GM12168* (FBgn0167108)	scaffold_4:4874849..4877078
	D. yakuba	*GE23401* (FBgn0240592)	3R:26956123..26958286
	D. erecta	*GG11952* (FBgn0104247)	scaffold_4820:1973564..1975704
	D. ananassae	*GF16192* (FBgn0093214)	scaffold_13340:21056664..21059300
	D. pseudoobscura	*GA22002* (FBgn0081987)	2:1482261..1484415
	D. persimilis	*GL14059* (FBgn0151664)	scaffold_7:4285074..4287260
	D. willistoni	*GK13141* (FBgn0215150)	scf2_1100000004943:9484681..9487579
	D. mojavensis	*GI24323* (FBgn0147046)	scaffold_6540:10097851..10101121

	D. virilis	*GJ10408* (FBgn0197690)	scaffold_12855:3095571..3099015
	D. grimshawi	*GH18422* (FBgn0125890)	scaffold_14906:5797460..5801093
CG9747	*D. melanogaster*	*CG9747* (FBgn0039754)	3R:26011566..26016463
	D. simulans	*GD17234* (FBgn0188797)	3R:25679027..25683922
	D. sechellia	*GM12172* (FBgn0167112)	scaffold_4:4864314..4869213
	D. yakuba	*GE23403* (FBgn0240594)	3R:26945103..26950087
	D. erecta	*GG11954* (FBgn0104249)	scaffold_4820:1981496..1986547
	D. ananassae	*GF16194* (FBgn0093216)	scaffold_13340:21045994..21051062
	D. pseudoobscura	*GA22005* (FBgn0081990)	2:1490305..1494712
	D. persimilis	*GL14061* (FBgn0151666)	scaffold_7:4293137..4297511
	D. willistoni	*GK13143* (FBgn0215152)	scf2_1100000004943:9472854..9477605
	D. mojavensis	*GI24325* (FBgn0147048)	scaffold_6540:10111520..10118082
	D. virilis	*GJ10410* (FBgn0197692)	scaffold_12855:3079765..3086153
	D. grimshawi	*GH18424* (FBgn0125892)	scaffold_14906:5783828..5789396
CG15531	*D. melanogaster*	*CG15531* (FBgn0039755)	3R:26020605..26021890
	D. simulans	*GD17223* (FBgn0188786)	3R:25688409..25689688
	D. sechellia	*GM12171* (FBgn0167111)	scaffold_4:4873106..4874385
	D. yakuba	*GE23402* (FBgn0240593)	3R:26954354..26955635
	D. erecta	*GG11953* (FBgn0104248)	scaffold_4820:1976171..1977465
	D. ananassae	*GF16193* (FBgn0093215)	scaffold_13340:21054974..21056242
	D. pseudoobscura	*GA26941* (FBgn0248312)	2:1484823..1486141
	D. persimilis	*GL14060* (FBgn0151665)	scaffold_7:4287668..4288986

D. willistoni	*GK13142* (FBgn0215151)	scf2_1100000004943:9482984..9484270
D. mojavensis	*GI24324* (FBgn0147047)	scaffold_6540:10102110..10103733
D. virilis	*GJ10409* (FBgn0197691)	scaffold_12855:3093065..3095092
D. grimshawi	*GH18423* (FBgn0125891)	scaffold_14906:5795284..5796956

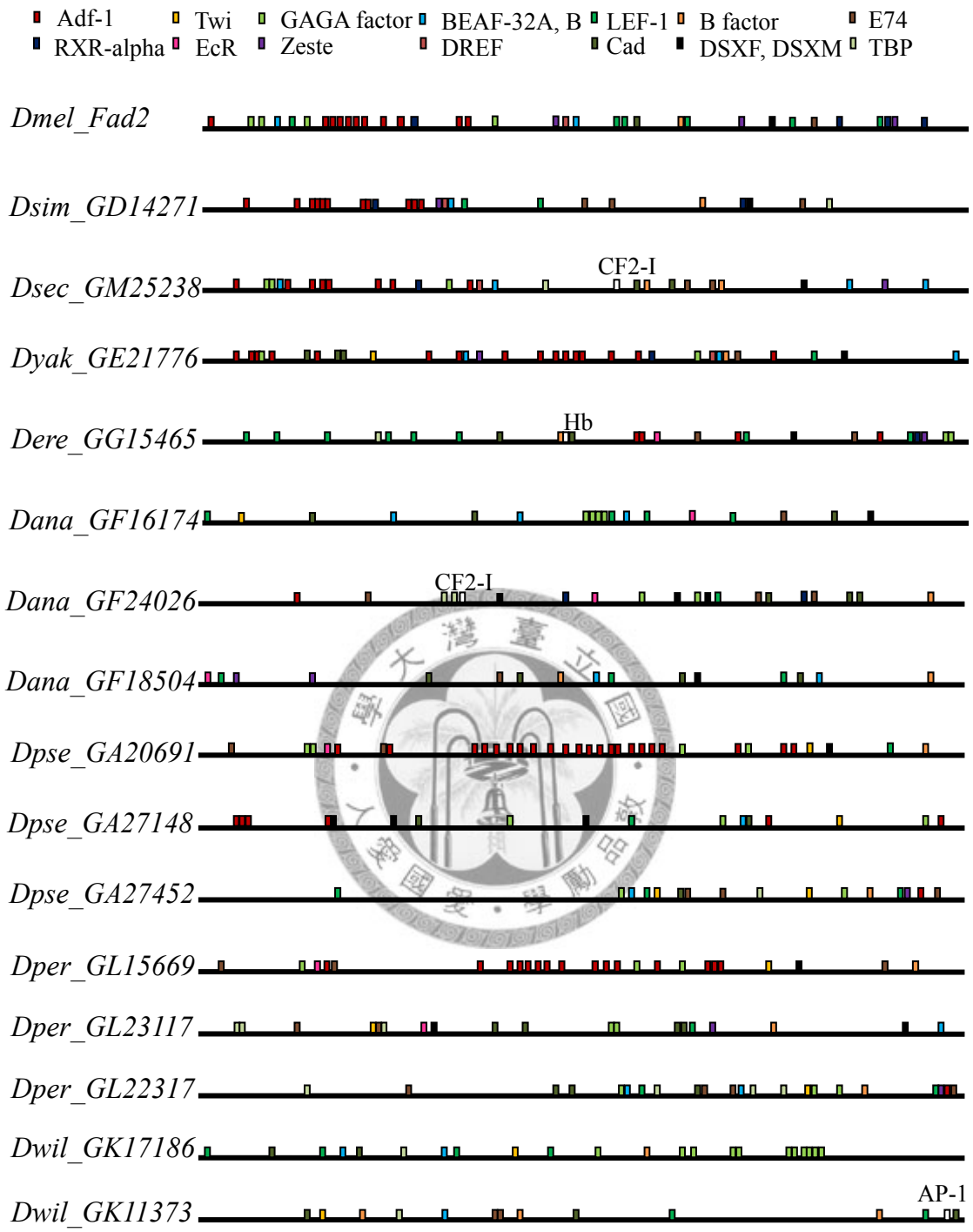


Fig S1: The annotated transcription binding sites on the 5' flanking regions of *Fad2* homologs. The sites highly dispersed among species are shown as color boxes and the others are show in white ones.

Table S3: The list of annotated transcription factors on the 5' flanking regions of *Fad2* homologs.

Transcription binding sites	Sequence	Function
Alcohol dehydrogenase transcription factor (Adf-1)	GCTGC	It is related with the transcription of Alcohol dehydrogenase (England, Heberlein, and Tjian 1990)
AP-1 (*Antennapedia* protein 1)	TGATTCA	It was involved in the regulation of *Antennapedia* expression during *Drosophila* embryo development (Perkins, Dailey, and Tjian 1988).
Boundary element-associated factor of 32 kDa (BEAf-32A, BEAf-32B)	CGATA	Boundary elements were associated with gene expression by limiting possible interactions between regulatory elements and promoters (Kuhn and Geyer 2003).
B factor	TATAAAA	It was required for *Drosophila* RNA polymerase II to initiate the transcription (Parker and Topol 1984).
Caudal (Cad)	TTTAGG	It would form an anterior-to-posterior concentration gradient during *Drosophila* embryo development opposite to gradient form by *bicoid* (Rivera-Pomar et al. 1995).
CF2-I	TATATTATA	It was responsible for developmentally-regulated alternative splicing that makes genes produce different protein (Hsu et al. 1992).
DNA replication-related element binding factor (DREF)	TTTCGATA	It activates the transcription of genes related to DNA replication and cell proliferation (Sawado et al. 1998).
DSXF, DSXM	CTACAATGT	The binding of DSXF or DSXM would cause the sex-biased expression of genes (Coschigano and Wensink 1993).
E74	AGGAA	It was an acdysone-induced gene and E74A promoter was activated by ecdysone and was subsequently repressed by ecdysone-induced proteins in larval stage (Burtis et al. 1990; Thummel, Burtis, and Hogness 1990).

Ecdysone receptor (EcR) and Retinoid X receptor (RXR-alpha)	And TGCACT	They would form a heterodimer that were stabilized and activated by ecdysone (Thomas, Stunnenberg, and Stewart 1993).
GAGA factor	CTGTC	It was very important for regulating the transcription of numerous genes in *Drosophila* (Wilkins and Lis 1998).
Hunchback (Hb)	AGTAAAAAC	It was involved in the *Drosophila* body pattern formation during embryogenesis (Struhl, Johnston, and Lawrence 1992).
Lymphocyte enhancer binding factor 1 (LEF-1)	TTTGA	It was related with *armadillo* that affect the posterior segmentation and lead to the formation of anterior structures (Riggleman, Wieschaus, and Schedl 1989).
TATA binding protein (TBP)	TATTTAA	It was required for RNA polymerase activating (Klages and Strubin 1995).
Twist (Twi)	CACGTG	It plays an important role in the development of *Drosophila* embryo like the formation of mesoderm, dorsal-ventral pattern and larval cuticles (Nuesslein-Volhard, Wieschaus, and Kluding 1984; Thisse, el Messal, and Perrin-Schmitt 1987; Leptin 1991).
Zeste	CACTCA,	It would activate the *Ultrabithorax* transcription in the *Drosophila* embryo (Laney and Biggin 1992).

Table S4: Species and strains used in population genetics analysis.

Species	Strain	Source
D. simulans	W55	Toulon, France
	W56	Chapel Hill, USA
	W58	Australia
	W60	Rakujyuen, Mishima, Japan
	W61	Izumi, Kagoshima, Japan
	W66	St. Denis, Reunion
	W73	
	W74	
	W75	
	W79	Tananarive, Madagascar
	W81	
	W86	
	W90	Nairobi, Kenya
	W91	
	W92	
	W100	Ootuki, Japan
	W112	Zimbabwe
	W113	
	W115	
	W116	
D. yakuba	CY01	Cammeroon (Andolfatto 2002)
	CY02	
	CY03	
	CY04	
	CY07	
	CY08	
	CY09	
	CY12	
	CY13	
	CY16	
	CY17	
	CY18	
	CY19	

	CY20	
	CY21	
	CY22	
	CY23	
	CY24	
	CY26	
D. teissieri	CT01	Cameroon (Andolfatto 2002)
	CT02	
	GT53w	Gabon (Ballard/Charlat 2000)
	GT69w	
	TLop6	Lope Forest, central Gabon
	TBra8	Daniel Lachaise in Brazzaville, Congo.
	TBra11	
	TBra12	
	TBRZ8	
	TBRZ17	
	TUZ11	Udzungwa Mounts , Tanzania
	TUZ35	
	TUZ125	
	T128.2	From Manyuan Long
	TT103	
	TTR47	


國立臺灣大學
人愛國愛學勵品教