

# 修士学位論文

## 題名

Pooled RNA-seqを用いた

アカシヨウジョウバエ低温耐性候補遺伝子の探索

指導教授 田村 浩一郎 教授

平成 28年 1月 8日 提出

首都大学東京大学院

理工学研究科 生命科学専攻

学修番号 14881324

氏名 中村 遥

# Pooled RNA-seq を用いた アカシヨウジヨウバエ低温耐性候補遺伝子の探索

進化遺伝学研究室 中村 遥

アカシヨウジヨウバエ (*Drosophila albomicans*) は近年、分布を熱帯から温帯へ拡大させ、現在の分布北限は西日本から中部地方であると考えられる。当研究室で行われた先行研究によると、本種は近縁種よりも低温耐性が高く、特に温帯の集団は熱帯の集団よりも平均して高い低温耐性を持つことが明らかとなった。このことから、本種の分布拡大の背景には低温耐性獲得による適応進化の可能性が考えられる。そこで本研究では、アカシヨウジヨウバエの低温耐性に関与した遺伝子を pooled RNA-seq を用いて網羅的に探索し同定することを目的とした。

pooled RNA-seq は様々な系統の個体を混ぜて RNA 抽出を行い、RNA-seq によって集団全体の平均的遺伝子発現量やエキソンの単一塩基多型 (SNP: single nucleotide polymorphism) の情報を得る手法である。先行研究からアカシヨウジヨウバエの低温耐性は、一定期間、適度な低温にさらされる「低温順化」によって有意に向上することがわかっている。また呼吸量も低温順化によって向上する。そこで採集場所と採集時期によってアカシヨウジヨウバエを 3 集団 (東南アジア集団、1991 年西日本集団、2011 年西日本集団) に分け、低温順化を行った条件下で total RNA を抽出し、Illumina HiSeq 2000 を用いて pooled RNA-seq を行った。

得られた SNP 情報をもとに、各集団間で集団分化の指数である  $F_{st}$  を計算した。その結果、東南アジア集団と西日本集団の間では集団分化が大きく、1991 年と 2011 年の西日本集団間ではあまり遺伝的分化が起こっていないということがわかった。このことは、20 年間、西日本集団において集団の遺伝的構成はあまり変化しておらず、遺伝的構成を変える自然選択が働いた痕跡は見いだせなかった。また Tajima's  $D$  によってアカシヨウジヨウバエ西日本集団の動態を推定したところ、西日本への進出に伴い集団サイズが減少していることが示唆された。先行研究より台湾に低温耐性の高い系統が存在し、日本のアカシヨウジヨウバエは台湾由来であることが分かっていることから、高い低温耐性の原因遺伝子は台湾集団でスタンディングバリエーションとして存在し、そのような遺伝子を持つアカシヨウジヨウバエが西日本に進出したシナリオが考えられる。

一方、東南アジア集団と西日本集団の間では、22 遺伝子の発現量の差が統計的に有意に異なった。その中 2 遺伝子 (*Cyp12d1-d*, *CG114322*) が、Isobe (2014) において低温順化によって遺伝子発現量に変化する遺伝子に含まれた。さらに

統計的に有意ではないが発現量差がある遺伝子の中にも 1 遺伝子 (*CG11889*) が含まれた。それらを低温耐性関連候補遺伝子とし、それらに関して集団内の遺伝的多様性を調べ、低温環境への適応進化に伴う自然選択の有無を検証した。その結果、それらの遺伝子の多様性は東南アジア集団に比べ西日本集団では低くなっていることがわかり、自然選択の関与が示唆された。しかしこれらの遺伝子の機能と低温耐性との直接の関与は考えにくく、他の研究の結果も合わせ、アカショウジョウバエでは低温ストレスに対し複数のストレス応答シグナルがクロストークしている可能性が示唆された。

## 学位論文要旨 (修士 (理学))

論文著者名 中村 遥

### **Detection of genes for cold tolerance in *Drosophila albomicans* using pooled RNA-seq.**

In a recent few decades, *Drosophila albomicans* has expanded their distribution from tropical zone to temperate zone. It has been reported that the current northernmost limit of *D. albomicans* distribution is west or central Japan. In previous studies, the variation of cold tolerance among *D. albomicans* strains was observed: the temperate zone population has a stronger cold temperature than the tropical population. This suggests that *D. albomicans* has expanded their distribution to temperate zone by adapting to the cooler climate. Therefore, I tried to identify genes responsible for the adaptation, using pooled RNA-Seq.

The pooled RNA-Seq is a very challenging way by applying RNA Seq to a pool of many strains. Because it is very new, I should create pipeline for analyzing the data from a next generation sequencer. However, I could get information on gene expression and single nucleotide polymorphism (SNP) in exons at a population level in a cost-effective way. According to previous studies, it was found that cold tolerance of insect was enhanced by cold acclimation. It was reported that cold tolerance was enhanced by a cold acclimation at 20 °C for several days accompanied by changes in expression level of many genes in a strain of *D. albomicans*, but these changes have not been examined at a population level.

In this study, the total RNA was extracted from cold-acclimated flies from three populations classified by location and time, i.e., from Southeast Asia (SEA), Japan in 1991 (J-1991) and Japan in 2011 (J-2011). Using the transcriptome data obtained by pooled RNA-Seq by Illumina HiSeq2000 sequencer, I compared differences in gene expression levels and genetic variations at a genomic level among the populations.

I calculated  $F_{st}$ , which is a measure of population differentiation due to genetic structure. As the result, I found that the genetic structure was different between SEA and the J-1991 and J-2011 populations, whereas genetic structure has not been differentiated between J-1991 and J-2011, compared to those and SEA. This indicates that in the west Japan population the genetic structure has not been changed during

the last 20 years, suggesting that west Japan population has not been under strong natural selection to change the genetic structure of population. In addition, using Tajima's  $D$ , I estimated that population size has been decreased during the distribution expansion to west Japan.

I found that gene expression level was different in 22 genes between SEA and the west Japan population. Among these genes, three genes (*Cyp12d1-1*, *CG13422* and *CG11889*) were included in the genes whose expression level was changed by cold acclimation according to Isobe (2014). So those three genes are candidates of the genes to be related to cold tolerance. To examine the effect of natural selection for the adaptation to cold environment, I computed nucleotide diversity, Waterson's theta, and Tajima's  $D$  for those candidate genes. All parameters indicate that J-1991 and J-2011 have lower genetic diversity than SEA, suggesting some effect of natural selection. However, the expected molecular functions of those genes are unlikely involved in cold tolerance directly and the observed expression changes of those genes are more likely attributed to cross-talk with a signaling pathway of other genes responding to cold stress.

## 目次

序論	・・・	2
材料と方法	・・・	5
結果	・・・	12
考察	・・・	17
結論	・・・	25
謝辞	・・・	26
参考文献	・・・	27
表	・・・	36
図	・・・	39
附表	・・・	51
付図	・・・	57

## 序論

生物は様々な生息環境に適応し進化してきた (Darwin 1859)。その中で自然選択による形質の獲得は遺伝子構造の変化だけでなく、遺伝子発現量の変化による場合も存在し、その結果さまざまな表現型が生み出されてきた (King and Wilson 1975)。

さまざま生物種において、自然選択によって集団内の遺伝的構造に変化が起こってきたことが知られている。イトヨでは、海水棲集団と淡水棲集団で鱗板が大きく変わり、淡水棲集団では鱗板が形成されない。これは EDA 遺伝子の変化が原因となっていることが知られており、海水棲集団では低い頻度しか存在しない表現型が淡水棲集団では主要な表現型に変わっている (Colosimo et al. 2005)。これら表現型の違いは、EDA 遺伝子上流の転写調節領域における単一塩基多型 (SNP) が原因となり遺伝子発現量が変わることにより引き起こされ、自然選択によって集団内の頻度が変化したと考えられている (O’Brown et al. 2015)。しかし、このように遺伝子発現量の変化によって適応進化の遺伝的メカニズムが明らかにされた研究は少なく、特に自然選択による変化の過程を明らかにした例は少ない。そこで本研究では、分布を急速に拡大している生物を用い、適応進化の遺伝的メカニズムを明らかにすることを試みた。

テングショウジョウバエ亜群 (*Drosophila nasuta* subgroup) に属するアカショウジョウバエ (*Drosophila albomicans*) は、元々、東南アジアを中心とする熱帯に分布するショウジョウバエ種である (Kitagawa et al. 1982)。しかし 1980 年半ば以降、分布を北に拡大させてきており、現在では西日本や関東地方でも採集の報告がある (Fujino et al. 2006) (図 1)。このことは、アカショウジョウバエが分布域を熱帯から温帯へ急激に拡大させたことを示唆しており、温帯への適応進化が起こった可能性が考えられる。変温動物の分布北上には、冬の寒さに耐える低温耐性が大きく関与していることが知られている (Guschina and Harwood 2006 による総説)。そこで本研究では、採集地域、採集時期の

異なる複数の集団を用い、集団レベルの遺伝子発現量や塩基配列データの網羅的解析により低温耐性に関わる遺伝子の同定および自然選択による環境適応の検証を試みた。

アカショウジョウバエ種内において低温耐性の強い系統と弱い系統が存在し、温帯集団には低温耐性の強い系統の頻度が高いことがわかっている。また、低温耐性は、25 °Cで生育したハエが20 °Cに数日間さらされる低温順化により向上するということが明らかになっている (Isobe 2014)。低温順化は様々な昆虫でも報告されていることから (Fields et al. 1998、Baust and Miller 1972)、アカショウジョウバエにおいても低温順化は低温耐性に大きな役割を果たしていると考えられる。そこで低温順化したアカショウジョウバエと低温順化していないアカショウジョウバエについて、それぞれ Illumina GAIIx DNA シークエンサーを用いた RNA-Seq (Mortazavi et al. 2008) によるトランスクリプトーム解析が行われ、低温順化の有無によって発現量に変化する遺伝子が網羅的に調べられた (Isobe 2014)。しかしこの研究では、台湾で採集された1系統における遺伝子発現量の変化が調べられているが、この変化が集団レベルで共通するかどうかは調べられていない。

本研究では、pooled RNA-Seq (Duitama et al. 2012) と呼ばれる複数の系統のサンプルを混ぜて RNA-Seq を行う方法を用い、地域集団間のトランスクリプトームの平均的差異を調べた。この方法は、通常の RNA-Seq と同様ゲノム情報を必要としないため、非モデル生物においても効率よく遺伝子発現量およびエクソンでの配列変異情報を入手することができる (Miller 2013)。しかし、この方法はまだ新しい方法であるため、解析方法についても検討を行った。東南アジアおよび西日本で採集された系統を採集地域、採集時期で3集団に別け、低温順化後の各集団間における遺伝子発現量の差を調べることでアカショウジョウバエ低温耐性に関与する遺伝子の同定を試みた。Illumina HiSeq2000 シークエンサーを用いた pooled RNA-Seq により得られた配列データを使用し、集団間で発現量の異なる遺伝子を調べ、それら候補遺伝子の中から低温順化の有

無により発現量に変化する遺伝子と共通する遺伝子を探索し、低温耐性関連候補遺伝子とした。それらの遺伝子に関して塩基性多様度を求め自然選択の影響を検証した。また SNP 情報から Tajima's  $D$  (Tajima 1989a, b)を推定し、アカショウジョウバエ集団における集団動態を予測した。

## 材料と方法

### アカショウジョウバエ系統

テングショウジョウバエ亜群 (*Drosophila nasuta* subgroup) に属するアカショウジョウバエ 60 単一雌由来系統を用いた。各系統は、採集地、採集時期によって 3 集団に分けた。具体的には、生息範囲の分布拡大以前に東南アジアで採集された「東南アジア集団 (SEA)」、分布拡大後、1990~1991 年に西日本で採集された「1991 年西日本集団(J-1991)」、および 2011 年に西日本で採集された「2011 年西日本集団(J-2011)」である。それぞれの集団は 20 系統からなる。本研究に用いたアカショウジョウバエの系統名と採集地点は図 2abc (詳細は付表 1) に示す。これら全てのショウジョウバエ系統は、標準コーンミール培地 (寒天・エビオス・粉末トウモロコシ・グルコース・プロピオン酸・ボーキニンからなる) の入った飼育ビン (直径 26 mm、高さ 103 mm) を使い、25 °C で飼育した。pooled RNA-Seq に使用したアカショウジョウバエは全灯条件で飼育した。

### AGPC シリカゲル法による RNA 抽出

各集団から羽化後、1 週間の雄 1 個体ずつ新しい飼育ビンに移し、低温順化を行ったものを用意し RNA の抽出を AGPC 法 (Chomczynski and Sacchi 1987) とシリカゲル法 (Boom et al. 1990) を合わせた方法で行った。

Solution D (4 M Guanidine Thiocyanate; 0.75 M Sodium Citrate; 0.5 % Sodium lauryl sarcosinate) 500  $\mu$ l を入れた 1.5 ml チューブの中に各集団 20 個体 (各系統 1 個体) の雄個体を入れ、バイオマッシャーII (フナコシ) を用いてよくすり潰した。その破砕液を 12,000 rpm、20 °C、1 分間遠心した後、上精をシリカゲル懸濁液 [高速液体クロマトグラフ用シリカゲル、直径 5  $\mu$ l、球状。0.01N 塩酸懸濁 (シリカゲル : 0.01N 塩酸 = 1 : 1)] 10  $\mu$ l が入った新しいチューブに移し、よく懸濁した後、1 分ごとにボルテックスを行い

ながら、氷上で5分間整地し、DNAをシリカゲルに吸着させた。その後12,000 rpm、20 °C、1分間遠心し、DNAを吸着したシリカゲルを沈殿させ、上清を別のチューブに移した。上清に2M Sodium Acetate (pH 4.0) 50 µl, Water-saturated Phenol 500 µl, Chloroform-Isoamylalcohol (24:1) 100 µlを加え、ボルテックスを行い、氷上で15分間静置した。その後4 °C、20分間遠心し、上層を回収した。回収した液量に対して半分量の Ethanol、シリカゲル懸濁液 10 µlを加え、よく混合した後、1分ごとにボルテックスしながら5分間静置し、RNAをシリカゲルに吸着させた。その後12,000 rpm、20 °C、1分間遠心し、上清をアスピレーターで除去した。沈殿したシリカゲルに Wash Buffer (10 mM Tris-HCl (pH 7.5); 100 mM NaCl : Ethanol = 1:4) 500 µlを加え、ボルテックスにて混合、12,000 rpm、20 °C、30秒間遠心、上清をアスピレーターで除去した。この操作をもう2回繰り返す、Guanidine Thiocyanateの洗浄除去を行った。沈殿物を風乾し残留 Ethanolを除去し、TE (10 mM Tris-HCl (pH 8.0); 0.1 mM EDTA) 50 µlを加え、ボルテックスで混合し、70 °Cで5分間インキュベートした後、12,000 rpm、20 °C、30秒遠心し、溶出したRNAを含む上清を滅菌済みチューブに回収した。

#### 次世代シーケンサーを用いた pooled RNA-Seq

次世代シーケンサーを用いた pooled RNA-Seqは、2012年12月にタカラバイオ株式会社によって行われた。タカラバイオ株式会社「東南アジア集団」「1991年西日本集団」「2011年西日本集団」のアカショウジョウバエから抽出した total RNAを送り、品質検定、シーケンスライブラリー作成 (Agilent 社) および HiSeq システム (イルミナ社) による高速シーケンス解析が実施された。以下は作業報告書の内容を記す。

## 品質検定

受け入れた検体に対して Nanodrop および Agilent 2100 Bioanalyzer を用いた品質管理が行われ、濃度および純度が評価された。

## アガロースゲル電気泳動

電気泳動には吸光値から約 100 ng が用いられた。

## 品質検定

受け入れ検体の品質検定が行われた結果、サンプルに問題が無いと判断され、シーケンスライブラリーが作成された。しかし、作製不良となったため、これらの検体に対し QIAGEN RNeasy MinElute Cleanup Kit (QIAGEN 社)を使用した精製処理による total RNA の品質向上が試みられた。精製検体の品質検査が行われた結果、検体の品質向上が確認されたため、これらの検体が引き続きシーケンスライブラリー作製に用いられた。

## ライブラリー作製

品質検定において解析可能と判断されたサンプルを用いて、自動化装置 (Agilent 社) を使用しシーケンスライブラリーの作製が行われた。

## 解析プロトコール

ライブラリーの作製は、TruSeq RNA Sample Preparation V2 Kit (イルミナ社) を用い、TurSeq RNA Sample Preparation V2 Guide Rev. C マニュアル (イルミナ社) に従い行われた。解析検体より、PolyA+ RNA が単離され、断片化が行われ、この RNA 断片を鋳型として二本鎖 cDNA が合成された。合成された二本鎖 cDNA の両末端を平滑化・リン酸化処理した後、3'-dA 突出処理が行われ、Index 付きアダプターに連結された。アダプ

ターを連結した二本鎖 cDNA を鋳型とし、PCR による増幅が行われ、AMPureXP（ベクマン・コールター社）を用いた磁気ビーズ法にて得られた PCR 産物が精製され、シーケンスライブラリーとされた。

### **Agilent 2100 Bioanalyzer を用いた検定**

作製されたシーケンスライブラリーの品質は Agilent 2100 Bioanalyzer を用いて測定された。なおライブラリーはアダプターが付加されているため、ピークサイズからアダプターサイズ(約 100 bp)を除いたサイズがクローニングサイズとなっている。

### **シーケンス解析**

#### **解析プロトコール**

シーケンスライブラリーを用いて、シーケンスの鋳型となるクラスターが形成され、鋳型 DNA の塩基配列が取得された。シーケンスの解析は付属のソフトウェアを使用しベースコールが行われ、fastq 形式ファイルとして出力された。これらの作業は、シーケンス解析マニュアルを用いて行われた。

#### **シーケンス解析結果**

シーケンス解析により得られたリード数及び塩基数をまとめた。なお、DIL01ACXX-Lane3 については、Index 配列リードにおいて N を含むリードが多く確認されたため、得られた Index の配列リードに対し、1 塩基ミスマッチ（その他は完全マッチ）を許容し抽出したリードをライブラリー由来のリードとし、一致しなかったリードは Undetermined とされた。

## pooled RNA-Seq による transcriptome 解析および集団遺伝学的解析

### Quality control および *de novo* コンティグの作製

pooled RNA-Seq で得られた配列は fastq ファイルでのクオリティ値が 20 以上の配列のみを使用した。それらの配列の集団をすべて 1 つのファイルにまとめ、Trinity (Grabherr 2011) により *de novo* アセンブリを行い、*de novo* コンティグを作製した。得られた各 *de novo* コンティグをクエリーとし、既に全ゲノム配列が明らかになっているキイロショウジョウバエの全タンパク質コード配列データに対してローカル BLAST ソフトウェア (Shiryev et al. 2007) を用いてホモロジーサーチを行い、遺伝子のアノテーションを行った。アノテーションできたものを *de novo* リファレンス配列としてその後の操作に使用した。すなわちアカショウジョウバエの遺伝子名はすべてキイロショウジョウバエの遺伝子名として使われている名称に従った。

### マッピング、重複リードの除去、ファイルのソート、 $F_{st}$ の推定

PoPoolation2 (Kofler et al. 2011) のチュートリアルに従い以下の方法で集団分化の指数である  $F_{st}$  を推定した。

Bowtie2 (Langmead and Salzberg 2014) を用い、*de novo* リファレンス配列に対し pooled RNA-Seq で得られた各リードを集団ごとにマップした。得られた sam ファイルを SAMtools (Heng et al. 2009) によって bam ファイルに変換し、mapping quality が 20 以下のリードを除去した。その後 scaffold number および座標順で並び替えを行った。また RNA-Seq 特有の問題である遺伝子間の depth の違いを修正するため、マッピングされたリードとまったく同じ重複リードを除去した。それらのファイルを二つずつ合成し、mpileup ファイルとした。mpileup ファイルは PoPoolation2 を用い、sync ファイルとし

た。sync ファイルを PoPoolation2 によって、SNP ごとの  $F_{st}$  を集団ごとに計算した。また集団ごとの対立遺伝子頻度も計算した。

また R を用い、SNP ごとの  $F_{st}$  分布を Boxplot を使うことで箱ひげ図に表した。その際、SNP 総数が多いため、外れ値も多く、重複によって個々の点がつぶれてしまったため、外れ値を除外して箱ひげ図を作成した。

### 遺伝子配置の決定

*de novo* リファレンス配列をクエリーにし、アカショウジョウバエのドラフトゲノム (Nozawa, 未発表) に対してローカル BLAST ソフトウェアを用いてホモロジーサーチを行った。その結果からドラフトゲノム内の遺伝子の位置情報を得た。また Chang et al. (2008) によるアカショウジョウバエ染色体の Muller's element 構成に基づき (図 3)、ドラフトゲノムの Scaffold がどの Muller's element および染色体由来であるかを推定した。

### 集団ごとの発現量解析

集団ごとの遺伝子発現量は edgeR ver3.8.6 (Robinson et al. 2009) を用いて計算した。*de novo* リファレンス配列に対して Bowtie2 を用い集団ごとに得られた pooled RNA-Seq の各リードをマッピングした。異なる場所に 2 回以上マップされたリードがあった場合も除去をせずに解析に使用した。出力された sam ファイルは SAMtools を用いて bam ファイルに変換し、遺伝子ごとにソートしたのち、mapping quality 20 以上のリードのみを解析に用いた。その後 eXpress (Forster et al. 2013) を用いて *de novo* リファレンス配列にマップされたリード数を求めた。1991 年西日本集団と 2011 年西日本集団は 2 回の独立試行とし、リード数をもとに edgeR を使い、東南アジア集団と西日本集団の 2 集団間の遺伝子発現量差を定量した。発現量変化の  $P$ -value を FDR (False Discovery Rate) で補正し  $Q$ -value が 0.1 以下のものを集団間で発現量の差がある遺伝子とした。

## 集団内多様性の推定

GATK (McKenna et al. 2013) に含まれる haplotypcaller を用いて、*de novo* リファレンス配列にマップされたリードから単一塩基多型 (SNP) を同定し VCF ファイルで出力した。VCF ファイルから集団内多様性を比較した。方法は GATK の web ページにあるマニュアル(<https://www.broadinstitute.org/gatk/>)に準拠した。

全てのリードをまとめ作製した *de novo* リファレンス配列に 3 集団それぞれから得られたリードを Bowtie2 によりマッピングした。Picard (Wysoker et al. 2013) により重複リードを除去したのち GATK haplotypcaller を用いて SNP を同定した。その際 SNP の quality は q10 以上の変異を SNP として特定した。また挿入・欠失 (indel) は今回の解析では用いていない。各遺伝子内の SNP 数および集団内の頻度から塩基性多様度( $\pi$ : Nei and Tajima 1981) および塩基多型度( $\theta$ : Watterson 1975) を計算した。また解析には塩基置換に対する機能的制約が小さいと考えられるコドンの第 3 座位のみを用いた。その際、集団内の染色体数は 40 とし、それによって得られた頻度から塩基性多様度は計算した。 $\pi$  および  $\theta$  の計算には自作の java プログラムを使用した。pooled RNA-Seq 解析に用いた各プログラムのコマンドは付表 2 に記す。

## Tajima's $D'$ による集団動態の推定および自然選択の検出

GATK の haplotypcaller によって推定した  $\pi$  および  $\theta$  から Tajima's  $D$  を推定した。RNA-Seq による SNP 同定は集団内のレアアリアルを見逃しやすいという傾向があるため、本研究では Tajima's  $D'$  として定義している。また解析には塩基置換に対する機能的制約が小さいと考えられるコドンの第 3 座位のみを用い、アカショウジョウバエ各集団の動態を推定した。R を用い、集団ごとの Tajima's  $D'$  分布を boxplot で示した。

同様に pooled RNA-Seq によって集団間で発現量の差が見られた遺伝子についても Tajima's  $D'$  を推定し、自然選択の影響について考察した。

## 結果

### pooled RNA-Seq を用いたトランスクリプトーム解析

#### *de novo* リファレンス配列の構築と遺伝子同定

タカラバイオ株式会社に委託したシーケンス解析の結果、各集団 63,527,348～114,110,725 リード（各リード 100 bp）を取得した。塩基クオリティが q20 以上の部位のみを用い、*de novo* アセンブリを行い、集団ごとに *de novo* リファレンス配列を作製した。解析には Trinity によって出力された transcript contig を使用した（表 1）。これらの contig を blastx によってキイロショウジョウバエ (*D. melanogaster*) 遺伝子データベース (dmel-nui-AA-r5.48) に対して相同性検索を行い、遺伝子アノテーションを行った結果 8,540 遺伝子を同定した。

#### 遺伝子配置の決定

ドラフトゲノム配列 (Nozawa 未発表) に pooled RNA-Seq からアノテーションされた遺伝子の塩基配列をマッピングし、ドラフトゲノムの各 scaffold の染色体上の位置を予測した。その結果、各 scaffold に含まれるほぼ全ての遺伝子は、同一の Muller's element 由来ということがわかった。例えば、Scaffold001 にマップされた遺伝子の大部分が Muller's element C または D にマップされた。Chang et al. (2008) によると、Muller's element C および D はアカショウジョウバエの第 3 染色体に含まれるため、これらの遺伝子は第 3 染色体上に存在していると推定された。同様に、Scaffold002、Scaffold006、は X 染色体由来、Scaffold004、Scafflod005 は第 2 染色体右腕由来、Scaffold003、Scaffold007 は第 2 染色体左腕由来であると推定された。各 scaffold について推定された染色体上の位置は表 2 に示す。しかし、各 scaffold 上の遺伝子の並びをアカショウジョウバエとキイロショウジョウバエの間で比較したところシンテニーはほとんど見いだせず、それぞ

れの染色体における scaffold の位置関係については推測できなかった。

### **$F_{st}$ を用いた集団間分化の解析**

PoPoolation2 を用い、単一塩基多型 (SNP) の検出を行い、SNP ごとに集団間の  $F_{st}$  (Wright 1965) を求めた。その結果、276,421 箇所の SNP が同定され、それぞれについて  $F_{st}$  の値が得られた。それらの分布を箱ひげ図に示した (図 4)。その結果、1991 年西日本集団と 2011 年西日本集団の間の集団分化は、それらと東南アジア集団の間の分化に比べ小さいことがわかった。

### **Tajima's $D$ を用いた集団動態の推定**

pooled RNA-Seq によって配列が得られた遺伝子について、コドンの第 3 座位を用いて Tajima's  $D'$  を計算した。第 3 座位に SNP が見られなかった遺伝子は解析から除外した。Tajima's  $D'$  の分布を箱ひげ図によって集団ごとに示した (図 5)。その結果、1991 年西日本集団および 2011 年西日本集団は東南アジア集団に比べ Tajima's  $D'$  の分布が高い値に偏っていた。この結果は、アカショウジョウバエが西日本地方へ移入した際に集団サイズが減少した可能性を示す。

## 遺伝子発現量の網羅的解析による低温耐性関連遺伝子の探索

### Bowtie2 を用いたマッピング

Bowtie2 を用い、各リードを *de novo* リファレンス配列に対しマッピングし sam ファイルとして出力した。それらのマッピング結果は表 3 に示す。異なる場所に 1 回以上マップされたリードについては全ての場所にマップされたものとして除去せずに解析に使用した。得られた sam ファイルを bam ファイルに変換し、遺伝子ごとに sort を行った。得られたマッピングデータの内、SAMtools を用いて mapping quality が 20 以上のリードのみを発現量解析に用いた。

### edgeR を用いた発現量解析

上記「Bowtie2 を用いたマッピング」で得られたファイルについて、eXpress を用いてリード数のカウントを行った。その結果をもとに edgeR を用いて集団間の遺伝子発現量の比較を行った。 $F_{st}$  の解析によって 1991 年西日本集団と 2011 年西日本集団の間の集団分化は小さいことが分かったので、両集団をプールして東南アジア集団と西日本集団の 2 集団間で遺伝子発現量の比較を行った(図 6)。比較に用いる遺伝子数が非常に多いため、集団間遺伝子発現量差の *P*-value を BH 法 (Benjamini and Hochberg 1995) による FDR (False Discovery Rate) の調整を行った。本研究では  $FDR < 0.1$  の遺伝子を集団間遺伝子発現量差がある遺伝子とした。その結果、22 遺伝子の発現量が 2 集団間で有意に異なっていた (表 6)。得られた 22 遺伝子の中、低温順化の有無で発現量の変化が見られた遺伝子 (附表 3、Isobe 2014) と共通する遺伝子を調べた結果、*Cyp12d1-d*、*CG13422* の 2 遺伝子が見つかった。また、統計的に有意にはならなかったが、2 集団間で発現量差があり、また低温順化の有無で発現量の変化が見られた遺伝子として *CG11889* もあった。

## 低温耐性関連候補遺伝子に関する集団遺伝学的解析

### 集団内多様性の解析

得られた pooled RNA-Seq データから集団ごとに各遺伝子の塩基多型度( $\theta$ )、塩基多様度( $\pi$ )を計算した。それらの分布を箱ひげ図に示した (図 7)。また pooled RNA Seq による SNP の同定は次世代シーケンサー特有の問題でレアなアリルを発見しそこなうが、Sanger 法による SNP 頻度との相関はあることが分かっている (付図 1)。その結果、東南アジア集団に比べ 1991 年西日本集団と 2011 年西日本集団は  $\pi$ 、 $\theta$  ともに平均して低くなっていることがわかった。

そこで低温耐性関連候補遺伝子コドン第 3 座位を用いた  $\theta$  を比較することによって自然選択の検出を試みた。*cyp12d1-d* ( $\theta_{SEA}=0.026, \theta_{J1991}=0.025, \theta_{J2011}=0.009$ )、*CG13422* ( $\theta_{SEA}=0.008, \theta_{J1991}=0, \theta_{J2011}=0.004$ )、*CG11889* ( $\theta_{SEA}=0.012, \theta_{J1991}=0.003, \theta_{J2011}=0.003$ )となった。全遺伝子の平均 ( $\theta_{SEA}=0.014, \theta_{J1991}=0.007, \theta_{J2011}=0.007$ ) と比較した場合、東南アジア集団と 1991 年西日本集団、2011 年西日本集団の塩基多型度の差は大きく、西日本集団が低い傾向にあった (図 8)。

また同様にコドン第 3 座位を用い、塩基性多様度( $\pi$ )を計算したところ、*cyp12d1-d* :  $\pi_{SEA}=0.036, \pi_{J1991}=0.024, \pi_{J2011}=0.009$ 、*CG13422* :  $\pi_{SEA}=0.01, \pi_{J1991}=0, \pi_{J2011}=0.003$ 、*CG11889* :  $\pi_{SEA}=0.020, \pi_{J1991}=0.001, \pi_{J2011}=0.001$  となっていた。全遺伝子の平均 ( $\pi_{SEA}=0.016, \pi_{J1991}=0.010, \pi_{J2011}=0.009$ ) と比較すると、 $\theta$  と同じように東南アジアに比べ日本集団の  $\pi$  が低い傾向にあった (図 8)。

### 低温耐性関連候補遺伝子の Tajima's $D$

低温耐性関連候補遺伝子について Tajima's  $D'$  を求めたところ、*Cyp12d1-d* :  $D'_{SEA}=1.439, D'_{J1991}=-0.143, D'_{J2011}=0.355$ 、*CG13422* :  $D'_{SEA}=1.222, D'_{J1991}=0$  (SNP 無し)、 $D'_{J2011}$

= -0.727、CG11889 :  $D'_{SEA} = 2.012$ ,  $D'_{J1991} = -1.296$ ,  $D'_{J2011} = -1.296$  となった (図 9)。データが得られた全ての遺伝子を用いた場合、西日本集団は東南アジア集団に比べ Tajima's  $D'$  は高い傾向にあった (図 5) が、候補遺伝子に関しては西日本集団のほうが低い傾向にあった。このことは上記の低温耐性関連候補遺伝子は西日本集団において純化選択など負の自然選択の結果、多様性が減少した可能性を示唆している。

## 考察

### pooled RNA-Seq による非モデル生物における遺伝子の同定

本研究では、アカシヨウジョウバエを採集地、採集時期で3集団に分け、それぞれから20単一雌由来系統のハエを用いて pooled RNA-Seq を行い、低温耐性関連遺伝子の同定を試みた。アカシヨウジョウバエは非モデル生物であるため、全ゲノム配列データがまだ得られていない。そこで RNA-Seq で得られたリードを用いて *de novo* アセンブリを行い、遺伝子発現量の解析および exon の SNP 解析のための配列データを取得した。*de novo* アセンブリの結果、8,540 遺伝子がキイロシヨウジョウバエのデータベースから同定された。キイロシヨウジョウバエでは約 14,300 遺伝子が知られているため、少なくとも半数以上の遺伝子がキイロシヨウジョウバエとアカシヨウジョウバエの間で相同性があることがわかった。本研究では、mRNA は低温順化を行った成虫から抽出し、また RNA-Seq によって配列データを得たため、成虫期以外に特異的に発現している遺伝子や発現量が少ない遺伝子の配列は得られていない。そのため、実際にはもっと多くの遺伝子がキイロシヨウジョウバエとアカシヨウジョウバエの間で共有されていると考えられる。一方、RNA-Seq では、mRNA として転写されている遺伝子のみ配列決定されるため、少なくとも配列決定できた遺伝子は低温順化時に発現している遺伝子であることが分かる。

### ドラフトゲノムを用いたアカシヨウジョウバエ染色体構造の予測

アカシヨウジョウバエのドラフトゲノム配列 (Nozawa 未発表) を用いて得られた遺伝子の染色体上の位置を確定することを試みた。その結果、各 scaffold 上に見つかった遺伝子はほぼ単一の Muller's element 上にあることがわかった。しかし、それらの遺伝子の並び順に関しては、キイロシヨウジョウバエとの間に長いシンテニーは見つからず、各 scaffold が染色体上のどの位置に存在するかは同定することができなかった。特に各

scaffold 両端にはシンテニーがあまり見出せず、2~3 遺伝子ほどのシンテニーの場合も多かった。また隣り合う遺伝子間の距離が、キイロショウジョウバエでは数 Mb 離れている場合もあった。これは、アカショウジョウバエ種内において非常に多くの逆位多型が存在している(Clyde 1982)ことに由来しているかもしれない。

### **pooled RNA-Seq 法による塩基頻度の解析**

本研究では、pooled RNA-Seq データから遺伝子発現量およびエクソンの SNP 情報を得た。エクソンの SNP 情報を得る際、遺伝子間でマッピングされたリード数 (depth) の大きな差が見られた。既存の解析プログラムはゲノム DNA の pooled-Seq データを想定しているため、集団間で depth に大きな差がある領域は解析に使えない。また depth が浅い領域は、シーケンシングエラーと SNP を単純な頻度から求めることが困難である。そこで本研究の解析においては、マッピングされたリードの中に完全に同一なリードがある場合、それらは PCR による増幅産物であると判断し、重複を除去した。GATK などの SNP 解析ソフトにおいてもこの操作を求められることが多い。この操作を行うことにより集団間で depth 差が大きかった領域についても解析を行うことができた。また、SNP 解析に用いた PoPoolation2 は SNP サイトの判断に quality value を利用しないため、depth が浅いサイトでは SNP とシーケンシングエラーの区別が不正確で、そのようなサイトは解析から除外せざるを得なかった。そこで quality value も考慮に入れるより高度な SNP Caller であらかじめ SNP サイトを抽出し、PoPoolation2 ではそれらを全て使用するようになればより正確な計算ができると期待できる。

### **東南アジア集団と西日本集団の遺伝的分化**

集団分化の指標として  $F_{st}$  を用い、6,044 遺伝子の集団間の遺伝的分化を推定した。SNP ごとの  $F_{st}$  の分布を推定したところ、東南アジア集団と西日本集団においては集団分化

が認められ、1991年と2011年の西日本集団間では集団分化はあまり認められなかった。1991年西日本集団と2011年西日本集団では、採集された時期が約20年離れているが集団分化はあまり生じていない。一方、東南アジア集団は1980年代に採集された系統が中心で、それらと西日本集団を比較すると有意に分化していることがわかった。これはアカショウジョウバエでは集団構造は地理的要因に大きく影響されていることを示唆している。西日本集団においては20年近くの時間差があっても集団の遺伝的構成はあまり変わっていない。このことは西日本地域において、集団の遺伝的構成を変化させるような自然選択は働かなかったことを示唆しており、少なくとも正の選択により遺伝的構成が短時間で大きく変化した形跡は観察されなかった。

またドラフトゲノム配列の scaffold 内において、集団分化が顕著に生じている領域は検出されなかった。このことは二つの要因の影響によるものだと考えられる。一つはアカショウジョウバエ種内における逆位多型の影響である。低温耐性が高いアカショウジョウバエの系統に低温耐性が低い系統を低温による人為選択を行いながら6世代戻し交配した実験においては、逆位の影響で50%以上の遺伝子は低温耐性が低い系統由来のままであった(阿部 2012)。本研究においても、種内の逆位多型の影響により各系統間で相同組換えの抑制が起こり selective sweep の痕跡を発見することを困難にしている可能性がある。もう一つは RNA-Seq 特有の問題である。RNA-Seq では、RNA 抽出時に発現している遺伝子のエクソン領域のみの配列が決定される。そのため非転写領域の配列情報を得ることができない。よって制御領域の配列から selective sweep の痕跡を検出できる可能性は残されている。

## 低温耐性関連候補遺伝子の探索

東南アジア集団と西日本集団の遺伝子発現量を比べた結果、多数の遺伝子の発現量に集団間で差が見られた。それらを BH 法によって  $FDR < 0.1$  に調整した結果、22 遺伝子のみ有意な差となった。本研究においては、その中から Isobe (2014) により台湾系統において低温順化によって発現量が増えることが分かった遺伝子群を見つけることにより、低温耐性関連遺伝子を探索することを試みた。そのため、一般によく使われている  $FDR < 0.05$  よりも基準を甘く設定した。 $FDR < 0.05$  に設定した場合、集団間で発現量に差が見られた遺伝子数は 14 遺伝子となったが、その中で低温順化によって発現量が増える遺伝子は 2 遺伝子となった。 $FDR < 0.15$  に設定した場合、集団間で発現量の差が見られた遺伝子の数は 23 遺伝子に増えるが、その中で低温順化によって発現量が増える遺伝子は 3 遺伝子となった。このことはトランスクリプトーム解析において  $FDR$  の調整基準を細かく検討するよりも、他の実験やスクリーニング法と合わせることで候補遺伝子を絞ることの方が効果的であることを示唆する。

本研究では、低温耐性関連遺伝子のスクリーニング方法として、集団間の遺伝子発現量の差および低温順化による遺伝子発現量の差という 2 種類の実験を用いた。しかし、さまざまな研究 (Calderon 2009, Addo-Bediako 2002) において低温耐性との関係が示唆されている糖新生を含む代謝に関与する遺伝子は、集団間で発現量差のある遺伝子として発見できなかった。アカショウジョウバエの低温耐性は、低温順化による遺伝子発現量の変化が重要な役割を果たしていると考えられているが (Isobe et al. 2013)、そのような遺伝子は、本研究で調べた低温順化後の遺伝子発現量の集団間比較では検出されなかった可能性が考えられる。Isobe (2014) では、低温耐性関連候補遺伝子について、いくつかの系統間で低温順化後の遺伝子発現量と低温耐性の相関を調べた結果、それらの間にはほとんど相関は見いだせなかった。しかし、順化の有無による遺伝子発現量の差と低温耐性の差の間には強い相関関係が見出された。この結果は、低温耐性には複数の遺伝

子の発現が関与しており、個々の遺伝子の発現量よりも複数の遺伝子の発現量の組合せが重要であることを示唆していると考えられる。本研究でも低温順化の有無による遺伝子発現量の差を pooled RNA-Seq により集団レベルで網羅的に比較すれば、より有力な低温耐性関連候補遺伝子を発見できたかもしれない。

集団間で発現量に差が見られた遺伝子の中、2 遺伝子は低温順化によって発現量に変化する遺伝子の中に含まれていた。その内の 1 遺伝子、*Cyp12d1-d* は、主に DDT などの薬害耐性に関わる遺伝子として知られている (Festucci-Buselli et al. 2005)。もう一つの遺伝子、*CG13422* は、*GNBP-like3* と相同でグラム陰性菌に特異的に結合するタンパクをコードしている。この遺伝子は主に細菌などの感染によって発現が促進することが知られている (Gregorio et al. 2001)。どちらの遺伝子も先行研究においては、遺伝子発現量の変化と低温ストレス応答に関する直接的な関係は見つからなかった (Festucci-Buselli et al. 2005, Gregorio et al. 2001)。しかし、何らかのストレスに応答し発現量に変化する遺伝子なので (Festucci-Buselli et al. 2005, Gregorio et al. 2001)、アカシヨウジョウバエにおいてもストレスに対する応答として低温順化に応答した可能性が考えられる。またあるストレスに応答し、複数のストレス応答が起こる現象は「クロストーク」といわれ、あるストレスに応答して特定のシグナルパスウェイに関与する転写因子などの発現量が変わると、結果として複数のストレス応答が同時に引き起こされるということが知られている (Sinclair et al. 2013)。実際、キイロシヨウジョウバエにおいて低温ストレスと免疫システムはクロストークすることが知られている (Zhang et al. 2011, Bourg et al. 2008)。そこで本研究で見つかった遺伝子のシグナルパスウェイを詳しく調べれば、アカシヨウジョウバエの低温耐性に関係する転写因子などを発見し、低温耐性を司る遺伝子を発見できる可能性がある。

## Tajima's $D'$ を用いた集団動態の推定

pooled RNA-Seq を用いたアレル頻度の推定においては、次世代シーケンサー特有の高いエラー率により低頻度アレルとエラーの識別が困難で、低頻度アレルを発見しこなうことが報告されている (Konczal et al. 2014)。またシーケンスされる配列は mRNA 由来のものなので、アレル頻度はその発現量に依存する。そのため本研究で推定した Tajima's  $D$  はバイアスしていると考えられる。そこで本研究で得られた Tajima's  $D$  は、Tajima's  $D'$  と表記することにした。しかし、全ての集団について同じ条件で配列決定しているため、集団ごとの Tajima's  $D'$  は同じようにバイアスしており、その分布を集団間で比較することは意味があると考えられる。

Tajima's  $D'$  の分布を集団間で比較したところ、1991 年西日本集団、2011 年西日本集団では、東南アジア集団に比べ分布が正の方向にずれていた。この解析ではタンパクの機能的制約が小さいコドンの第 3 座位のみを使用しており、かつゲノムレベルでの比較を行っているため、Tajima's  $D'$  の分布には個々の遺伝子にかかる自然選択の影響より集団動態の影響が強く反映されると思われる (Tajima 1989b)。そのため、西日本集団において Tajima's  $D'$  の分布が正の方向にずれていたことは、アカショウジョウバエの分布拡大に伴い西日本集団の集団サイズが減少した可能性を示唆するものと考えられる。しかし Tajima's  $D'$  の変動には、集団サイズ以外にも様々な要因が影響すると考えられるため、西日本集団における集団サイズ以外の要因についても検証する必要がある。

## 環境変化に伴う遺伝的構造の変化

$F_{st}$  を指標として用いた集団間の遺伝的分化の解析において、東南アジア集団と西日本集団は遺伝的に分化していることがわかった。1991 年西日本集団は、アカショウジョウバエが生息域を拡大した当初の集団であると考えられる。このことは、1991 年西日本集団においては、自然選択以外にもボトルネック効果や創始者効果により東南アジア

集団との間に遺伝的分化が生じたことが予想される。よって東南アジア集団と 1991 年西日本集団だけを比較するだけでは低温耐性関連遺伝子を絞り込むことは難しいと考えられる。そこで 20 年後に採集された 2011 年西日本集団を比較対象に用い、西日本地域で継続的に働いてきた自然選択の効果を検出しようと試みた。

しかし東南アジア集団と 2011 年西日本集団の遺伝的分化を東南アジア集団と 1991 年西日本集団との遺伝的分化と比較したところ、それらはほとんど同程度であった。また、1991 年と 2011 年の西日本集団の遺伝的分化は、それらと東南アジア集団との遺伝的分化に比べてずっと小さいものであった。*pepck* 遺伝子の集団遺伝学的研究によると、現在も断続的に東南アジアから西日本地方へアカショウジョウバエの流入は起こっていると推測された (鳥居 2012)。本研究でのゲノム規模の集団間遺伝的分化の比較では、20 年の間、西日本集団の遺伝的構成はあまり変わらずに維持され、また東南アジア集団との遺伝的分化も維持されていることが分かった。この結果は、東南アジアから西日本地方への断続的な流入が起こっていないか、西日本集団において遺伝的構成を変化させない自然選択が働いているとことを示唆している。その中には染色体逆位や低温以外の様々な環境要因の影響など、ゲノムレベルでの進化に関与している可能性も考えられる。

アカショウジョウバエの西日本集団は台湾に由来することが知られている (Ohsako et al. 1994)。台湾のアカショウジョウバエには低温順化によって大きく向上し高い低温耐性を有するものがある (Isobe et al. 2013)。このことは台湾集団内には低温耐性のスタンディングバリエーションが存在しており、その中の低温耐性が強い個体が日本へ進出することで西日本集団を形成した可能性がある。スタンディングバリエーションに自然選択が働く場合、*selective sweep* の痕跡は観察されにくいことが報告されているため (Barrett and Schluter 2007)、本研究で *selective sweep* の痕跡が見つからなかったのも、低温耐性関連遺伝子が台湾集団にすでにスタンディングバリエーションとして存在して

いたためかもしれない。台湾集団にすでにスタンディングバリエーションとして存在しているかどうかは、今後、日本集団と台湾集団について低温耐性に関する表現型とゲノム情報を比較することで検証できると期待される。

## 結論

pooled RNA-Seq を用いた集団間トランスクリプトーム解析により発現量に集団間差が見られた遺伝子が複数見つかった。低温順化の有無により発現量に変化する遺伝子との比較から、2 遺伝子において集団間および低温順化の有無により発現が変化する低温耐性関連遺伝子であることが予測された。しかし、それらの遺伝子の機能は薬剤耐性や免疫に関連するもので低温耐性との直接の関連は考えにくい。これはアカショウジョウバエの低温耐性に関する遺伝機構が免疫やほかのストレス応答シグナルに関する遺伝機構とクロストークしているためと考えられる。本研究では、低温順化を行ったアカショウジョウバエの pooled RNA-Seq のみを行っているため、得られた候補遺伝子は実際に集団レベルで低温順化により発現量に変化する遺伝子かどうかはわからない。今後、低温順化を行っていないアカショウジョウバエについても pooled RNA-Seq を行い、比較することで、低温耐性関連遺伝子をより確実に発見できることが期待される。

一方、エクソン内 SNP 情報からこの 2 遺伝子はゲノムレベルと比較して西日本集団で多様性が低下しており、自然選択を受けた可能性を示唆している。またゲノム規模の SNP 情報からアカショウジョウバエの西日本進出に伴う集団動態を予測した結果、西日本集団では集団サイズが減少した可能性が示唆された。これらの結果、および西日本集団は台湾集団に由来することから、台湾集団でスタンディングバリエーションとして存在していた高い低温耐性を持つアカショウジョウバエが西日本へ進出し、現在の西日本集団を構築したことが考えられる。

#### 研究倫理

遺伝子組換え実験は首都大学東京の研究倫理委員会の承認を得て実施した。承認番号は26-13、27-12 である。

#### 利益相反

該当しない。

## 謝辞

本研究を行うにあたり、熱心かつ丁寧にご指導をいただいた田村浩一郎教授、高橋文准教授および瀬戸陽介博士、宮城竜太郎博士、田中健太郎博士に心より御礼申し上げます。また里村和浩さんをはじめとする研究室の方々にも感謝を申し上げます。また RNA-Seq 解析のプログラムに関しましては岩本榮介さんに大変お世話になりました。最後になりますが体育会自転車部の先輩後輩方には私生活において大変お世話になりました。

## 参考文献

Addo-Bediako, A., Chown, S. L., & Gaston, K. J. (2002). Metabolic cold adaptation in insects: a large - scale perspective. *Functional Ecology*, 16(3), 332-338.

Baust, J. G., & Miller, L. K. (1972). Influence of low temperature acclimation on cold hardiness in the beetle, *Pterostichus brevicornis*. *Journal of insect physiology*, 18(10), 1935-1947.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.

Boom, R. C. J. A., Sol, C. J., Salimans, M. M., Jansen, C. L., Wertheim-van Dillen, P. M., & Van der Noordaa, J. P. M. E. (1990). Rapid and simple method for purification of nucleic acids. *Journal of clinical microbiology*, 28(3), 495-503.

Barrett, R. D., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in ecology & evolution*, 23(1), 38-44.

Calderon, S., Holmstrup, M., Westh, P., & Overgaard, J. (2009). Dual roles of glucose in the freeze-tolerant earthworm *Dendrobaena octaedra*: cryoprotection and fuel for metabolism. *Journal of Experimental Biology*, 212(6), 859-866.

Chang, T. P., Tsai, T. H., & Chang, H. Y. (2008). Fusions of Muller's elements during chromosome evolution of *Drosophila albomicans*. *Zoological Studies*, 47(5), 574-584.

Chomczynski, P., & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical biochemistry*, *162*(1), 156-159.

Clyde, M. (1982). Inversion polymorphism in Southeast Asian populations of the *Drosophila nasuta* subgroup. *Genetica*, *58*(3), 161-175.

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D., & Kingsley, D. M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *science*, *307*(5717), 1928-1933.

De Gregorio, E., Spellman, P. T., Rubin, G. M., & Lemaitre, B. (2001). Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proceedings of the National Academy of Sciences*, *98*(22), 12590-12595.

Duitama, J., Srivastava, P. K., & Măndoiu, I. I. (2012). Towards accurate detection and genotyping of expressed variants from whole transcriptome sequencing data. *BMC genomics*, *13*(Suppl 2), S6.

Festucci - Buselli, R. A., Carvalho - Dias, A. S., Oliveira - Andrade, D., Caixeta - Nunes, C., Li, H. M., Stuart, J. J., Muir, W., Scharf, M. E., & Pittendrigh, B. R. (2005). Expression of Cyp6g1 and Cyp12d1 in DDT resistant and susceptible strains of *Drosophila melanogaster*. *Insect molecular biology*, *14*(1), 69-77.

Fields, P. G., Fleurat-Lessard, F., Lavenseau, L., Febvay, G., Peypelut, L., & Bonnot, G. (1998). The effect of cold acclimation and deacclimation on cold tolerance, trehalose and free amino acid levels in *Sitophilus granarius* and *Cryptolestes ferrugineus* (Coleoptera). *Journal of Insect Physiology*, 44(10), 955-965.

Forster, S. C., Finkel, A. M., Gould, J. A., & Hertzog, P. J. (2013). RNA-eXpress annotates novel transcript features in RNA-seq data. *Bioinformatics*, 29(6), 810-812.

Fujino, Y., Beppu, K., & Nakamura, H. (2006). Stratification of the *Drosophilid* assemblage in the research forest of AFC, Shinshu University. *Bulletin Shinshu University Alpine Field Center*, 4, 47-55.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Lin, F., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Palma, d. P., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N & Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, 29(7), 644-652.

Guschina, I. A., & Harwood, J. L. (2006). Mechanisms of temperature adaptation in poikilotherms. *Febs Letters*, 580(23), 5477-5483.

Isobe, K. (2014). Genetic analysis of cold tolerance in *Drosophila albomicans*. *PhD. dissertation in biological sciences at Tokyo Metropolitan University*.

Isobe, K., Takahashi, A., & Tamura, K. (2013). Cold tolerance and metabolic rate increased by cold acclimation in *Drosophila albomicans* from natural populations. *Genes & genetic systems*, 88(5), 289-300.

King, M. C., & Wilson, A. C. (2014). Evolution at two levels in humans and chimpanzees. *Essential Readings in Evolutionary Biology*, 188(4184), 301.

Kitagawa, O., Wakahama, K. I., Fuyama, Y., Shimada, Y., Takanashi, E., Hatsumi, M, Uwabo, M., & Mita, Y. (1982). Genetic studies of the *Drosophila nasuta* subgroup, with notes on distribution and morphology. *Japanese journal of genetics*, 57(2), 113-141.

Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, 27(24), 3435-3436.

Konczal, M., Koteja, P., Stuglik, M. T., Radwan, J., & Babik, W. (2014). Accuracy of allele frequency estimation using pooled RNA - Seq. *Molecular ecology resources*, 14(2), 381-392.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4), 357-359.

Le Bourg, É., Massou, I., & Gobert, V. (2009). Cold stress increases resistance to fungal infection throughout life in *Drosophila melanogaster*. *Biogerontology*, 10(5), 613-625.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Proceeding Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.

Miller, A. C., Obholzer, N. D., Shah, A. N., Megason, S. G., & Moens, C. B. (2013). RNA-seq-based mapping and candidate identification of mutations from forward genetic screens. *Genome research*, 23(4), 679-686.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, 5(7), 621-628.

Nei, M., & Tajima, F. (1981). DNA polymorphism detectable by restriction endonucleases. *Genetics*, 97(1), 145-163.

O'Brown, N. M., Summers, B. R., Jones, F. C., Brady, S. D., & Kingsley, D. M. (2015). A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene *EDA*. *eLife*, 4, e05290.

Ohsako, T., Aotsuka, T., & Kitagawa, O. (1994). The origins of the Japanese mainland population of *Drosophila albomicans*. *Japanese journal of genetics*, 69(2), 183-194.

Wysoker, A., Tibbetts, K., & Fennell, T. (2013) <http://picard.sourceforge.net>

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for

differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140.

Shiryev, S. A., Papadopoulos, J. S., Schäffer, A. A., & Agarwala, R. (2007). Improved BLAST searches using longer words for protein seeding. *Bioinformatics*, 23(21), 2949-2951.

Sinclair, B. J., Ferguson, L. V., Salehipour-shirazi, G., & MacMillan, H. A. (2013). Cross-tolerance and cross-talk in the cold: relating low temperatures to desiccation and immune stress in insects. *Integrative and comparative biology*, ict004.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585-595.

Tajima, F. (1989). The effect of change in population size on DNA polymorphism. *Genetics*, 123(3), 597-601.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical population biology*, 7(2), 256-276.

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, 395-420.

Zhang, J., Marshall, K. E., Westwood, J. T., Clark, M. S., & Sinclair, B. J. (2011). Divergent transcriptomic responses to repeated and single cold exposures in *Drosophila melanogaster*. *The Journal of experimental biology*, 214(23), 4021-4029.

阿部 (2012) アカショウジョウバエ低温耐性遺伝子の染色体マッピング. 首都大学東京  
修士学位論文

鳥居 (2013) 集団遺伝学的手法を用いたアカショウジョウバエ温帯適応の検出. 首都大  
学東京 修士学位論文

表 1 Trinity の出力結果

Counts of transcripts, etc	
Total no. of trinity 'genes'	41,235
Total no.of trinity transcripts	51,310
Percent GC	43.68
Stats based on ALL transcript contigs	
Contig N10	2,524
Contig N20	1,795
Contig N30	1,379
Contig N40	1,070
Contig N50	836
Median contig length	394
Average contig length	614.02
Total assembled bases	31,505,448
Stats based on ONLY LONGEST ISOFORM per 'GENE'	
Contig N10	2,310
Contig N20	1,644
Contig N30	1,263
Contig N40	969
Contig N50	753
Median contig length	371
Average contig length	572.66
Total assembled bases	23,613,640

表 2 Mullar's element に基づく各 scaffold の染色体上の位置予測

scaffold number	染色体名	Muller's element
Scaffold001	Neo-sex 染色体 (第 3 染色体)	C or D
Scaffold002	Neo-sex 染色体 (X 染色体)	A
Scaffold003	第 2 染色体右腕	E
Scaffold004	第 2 染色体左腕	B
Scaffold005	第 2 染色体左腕	B
Scaffold006	Neo-sex 染色体 (X 染色体)	A
Scaffold007	第 2 染色体右腕	E
Scaffold008	第 2 染色体左腕	B
Scaffold009	第 2 染色体右腕	E
Scaffold010	第 2 染色体右腕	E
Scaffold011	第 2 染色体左腕	B
Scaffold012	不明	不明
Scaffold013	Neo-sex 染色体 (第 3 染色体)	C or D
Scaffold014	第 2 染色体左腕	B
Scaffold015	第 2 染色体右腕	E

Scaffold015 より小さいサイズの scaffold は遺伝子数が少なく、染色体位置の予測は出来なかった。

表 3a 東南アジア集団に比べ西日本集団で遺伝子発現量が増加していた遺伝子

Gene name	Fold Change	Molecular Function
<i>CG14300</i>	119.4	chitin binding
<i>CG13422</i>	26.0	carbohydrate binding
<i>Lcp65Ac</i>	48.5	structural constituent of chitin-based cuticle
<i>Cyp6d5</i>	17.1	electron carrier activity
<i>Cyp12d1-d</i>	12.1	electron carrier activity
<i>CG34291</i>	9.2	unknown
<i>CG5999</i>	8.0	glucuronosyltransferase activity
<i>CG9360</i>	7.5	oxidoreductase activity
<i>CG14245</i>	8.0	chitin binding
<i>CG6188</i>	7.0	glycine N-methyltransferase activity
<i>GstD1</i>	7.0	glutathione transferase activity
<i>RanBPM</i>	8.0	Ran GTPase binding
<i>CG14565</i>	6.5	unknown

表 3b 西日本集団に比べ東南アジア集団で遺伝子発現量が増加していた遺伝子

Gene name	Fold Change	Molecular Function
<i>CG10814</i>	17.1	gamma-butyrobetaine dioxygenase activity
<i>CG42329</i>	22.6	transferase activity
<i>CG3259</i>	10.6	microtubule binding
<i>GstE8</i>	12.1	glutathione transferase activity
<i>CG31105</i>	10.6	sodium channel activity
<i>CG4998</i>	4.9	serine-type endopeptidase activity
<i>TrpA1</i>	4.6	ligand-gated ion channel activity
<i>Lac</i>	4.3	protein homodimerization activity
<i>CG3502</i>	9.2	aminopeptidase activity;

集団間で遺伝子コンテイングにマップされたリード数に差があった遺伝子の FDR を調整し、集団間遺伝子発現量差がある遺伝子として定義した。分子的機能は FlyBase から引用した。

表 4 Bowtie2 によるマッピング結果

	東南アジア集団	1991 年西日本集団	2011 年西日本集団
Total read	73,845,013	73,565,304	61,372,121
Aligned 0 times	32,867,248	36,870,700	28,612,366
Aligned exactly 1time	40,116,988	36,139,349	32,199,921
Aligned >1 times	860,777	555,255	559,834

複数の遺伝子にマップされたリードは重複してカウントした。

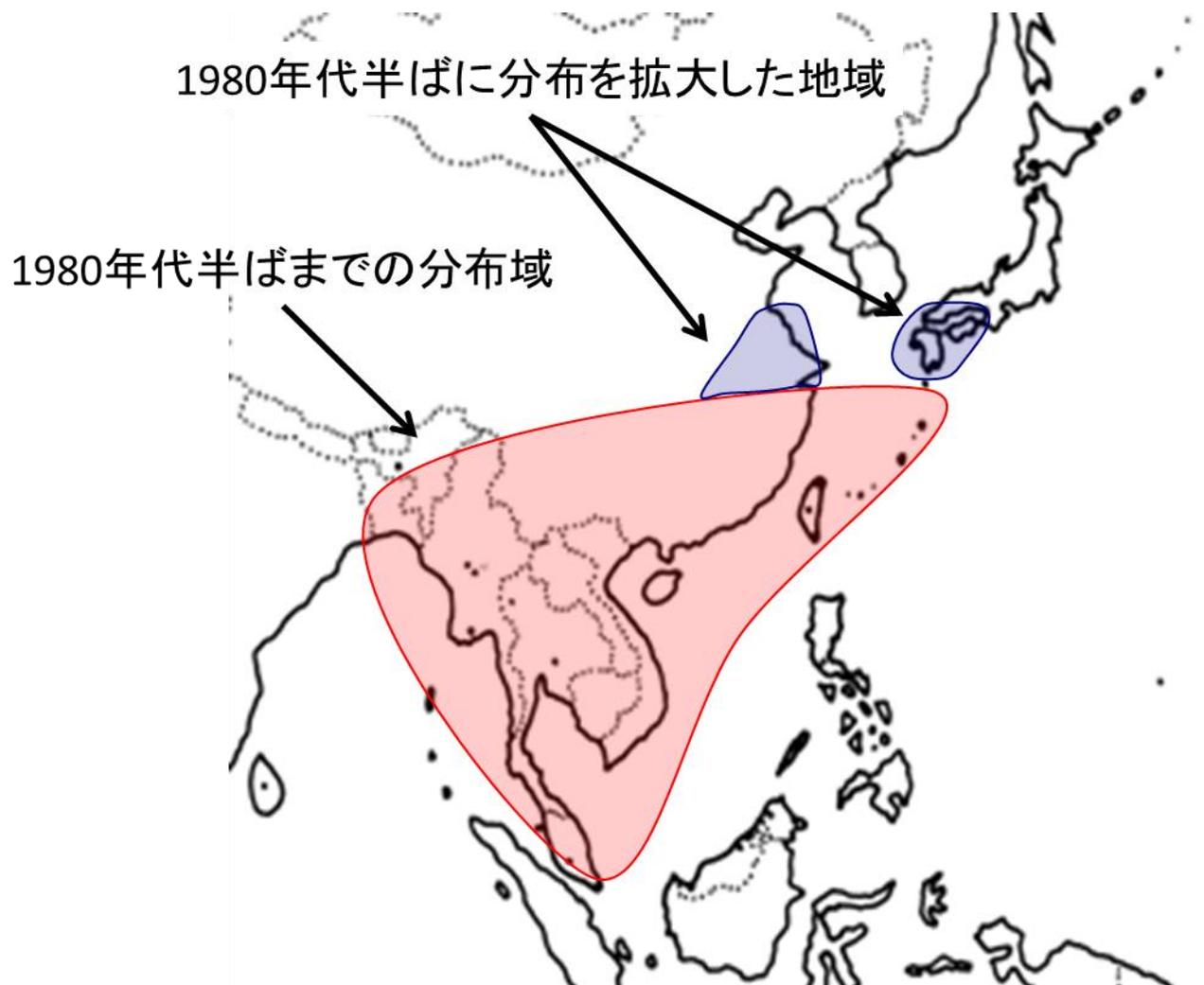


図1 アカシウジョウバエの分布域

アカシウジョウバエは元来、熱帯を中心(赤色)とする東南アジアに分布していた。しかし1980年代半ば以降西日本や中国沿岸(青色)の地域にも分布域を拡大させた。

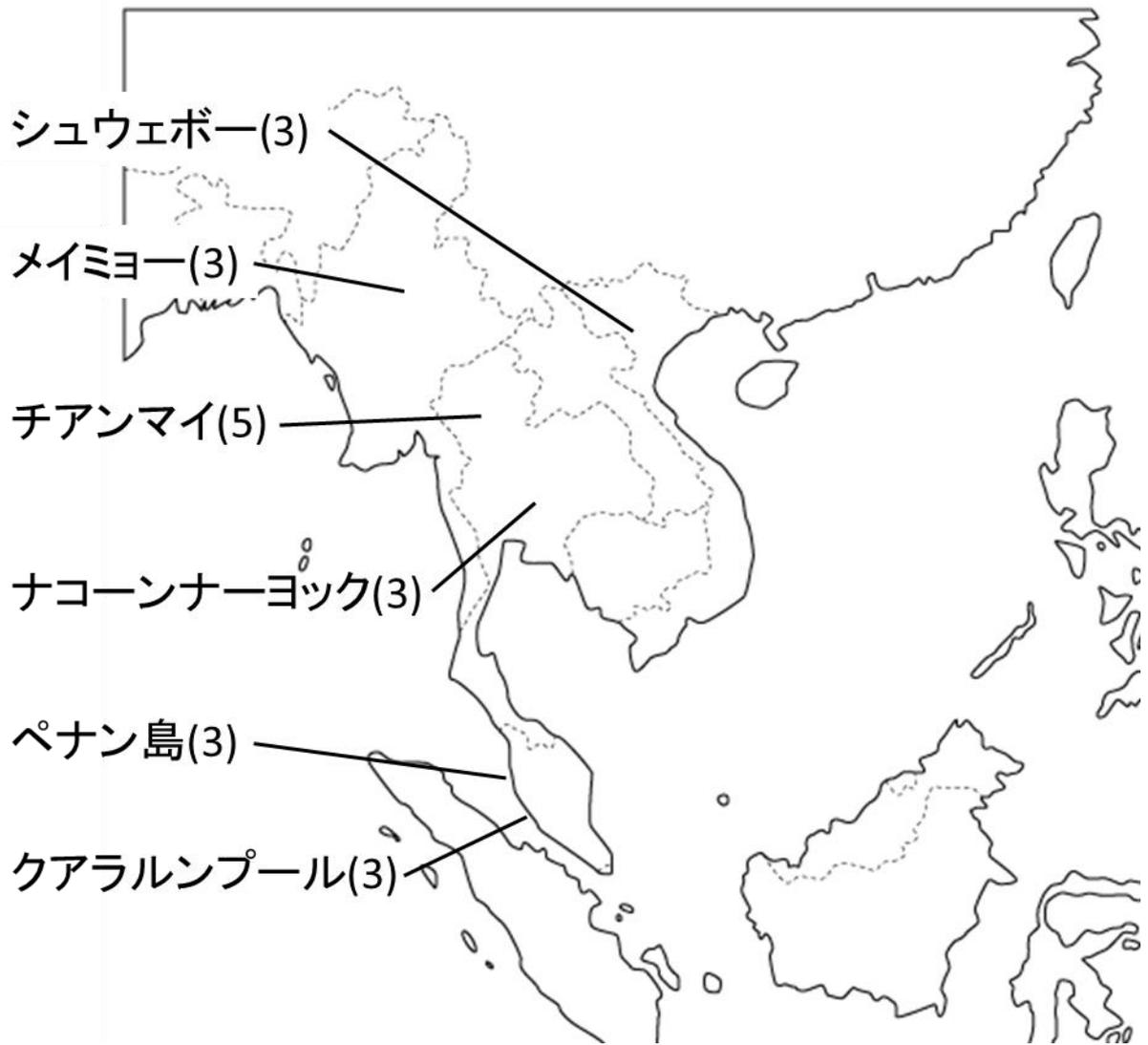


図 2a pooled RNA-Seq に用いたアカショウジョウバエの採集地点(東南アジア集団, SEA)  
 括弧内の数字は pooled RNA-Seq に使用した系統数を示す。

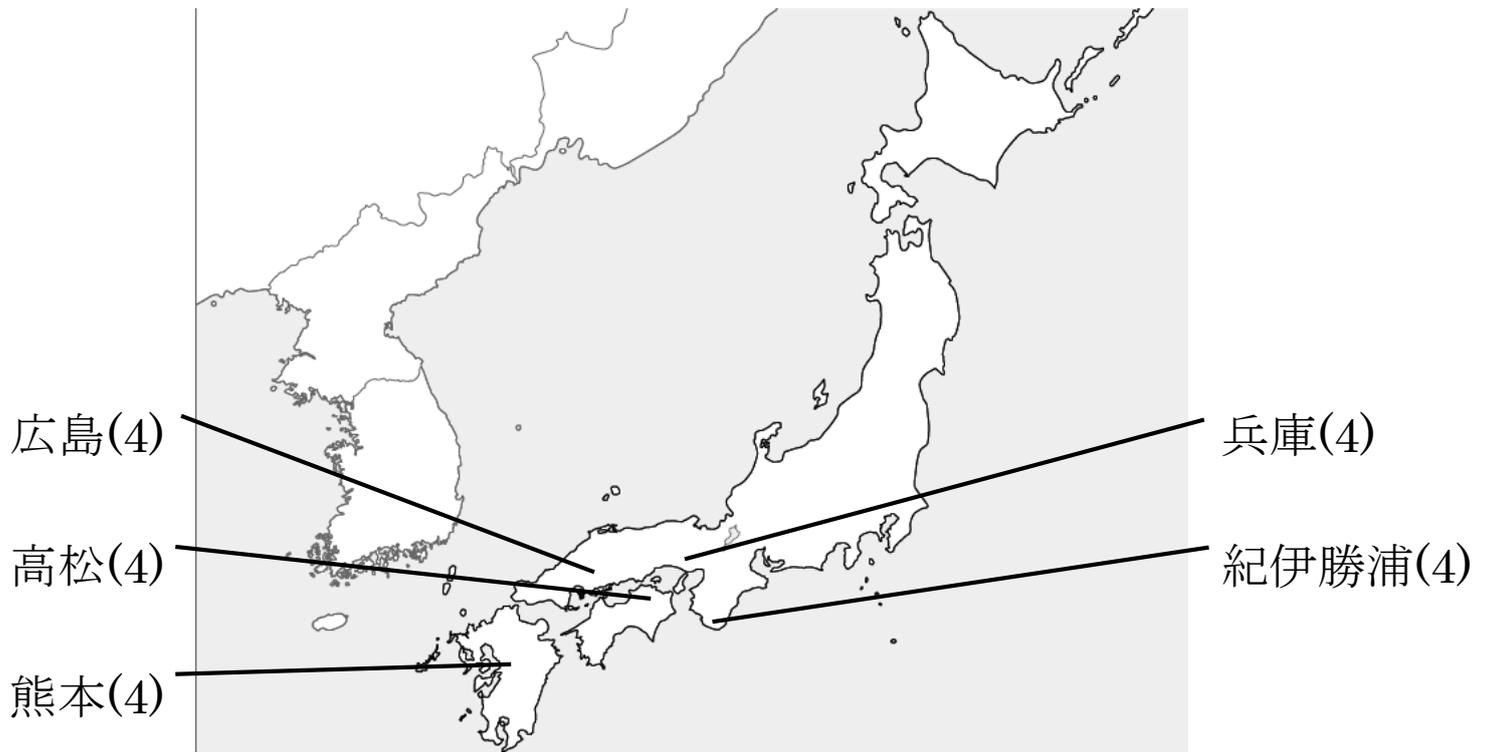


表 2b pooled RNA-Seq に用いたアカショウジョウバエの採集地点(1991 年西日本集団, J-1991)

括弧内の数字は pooled RNA-Seq に使用した系統数を示す。

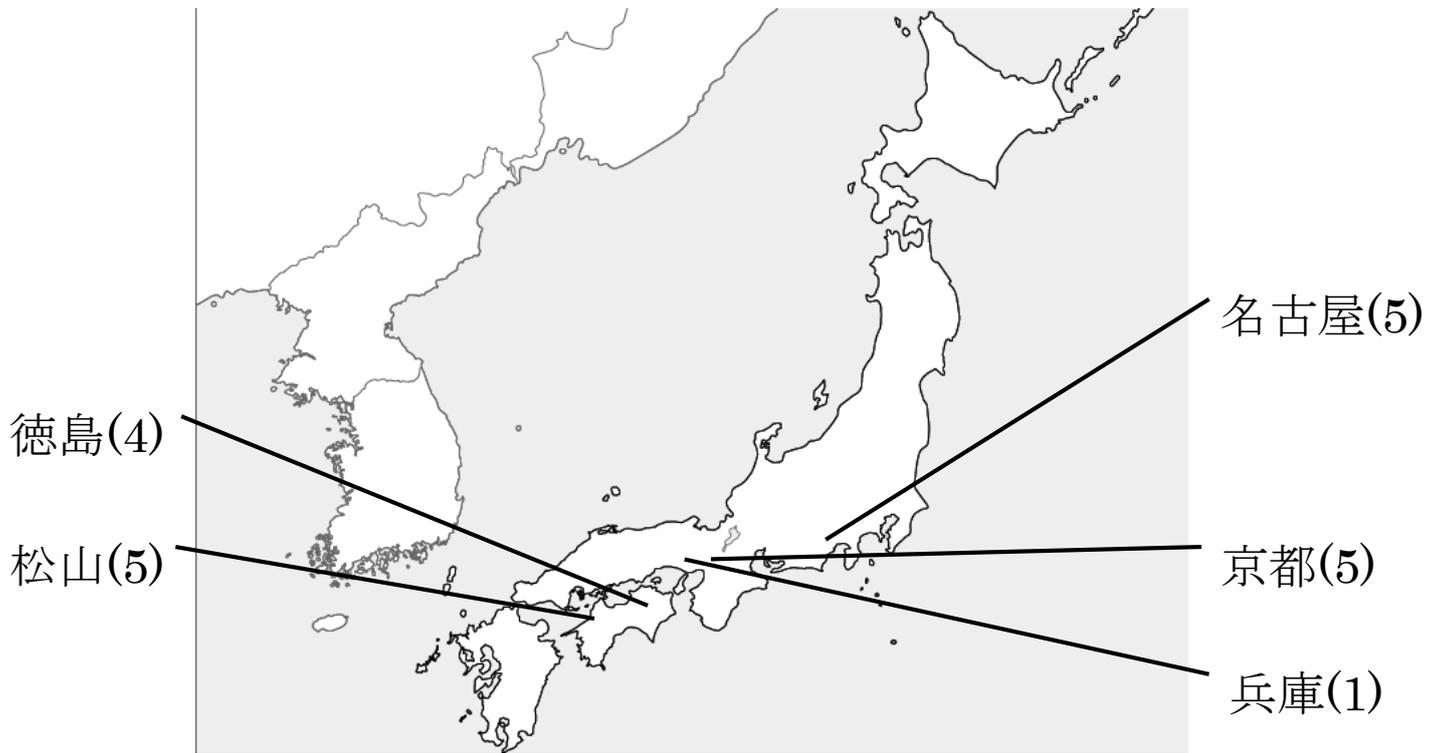


表 2c pooled RNA-Seq に用いたアカショウジョウバエの採集地点(2011 年西日本集団, J-2011)

括弧内の数字は pooled RNA-Seq に使用した系統数を示す。

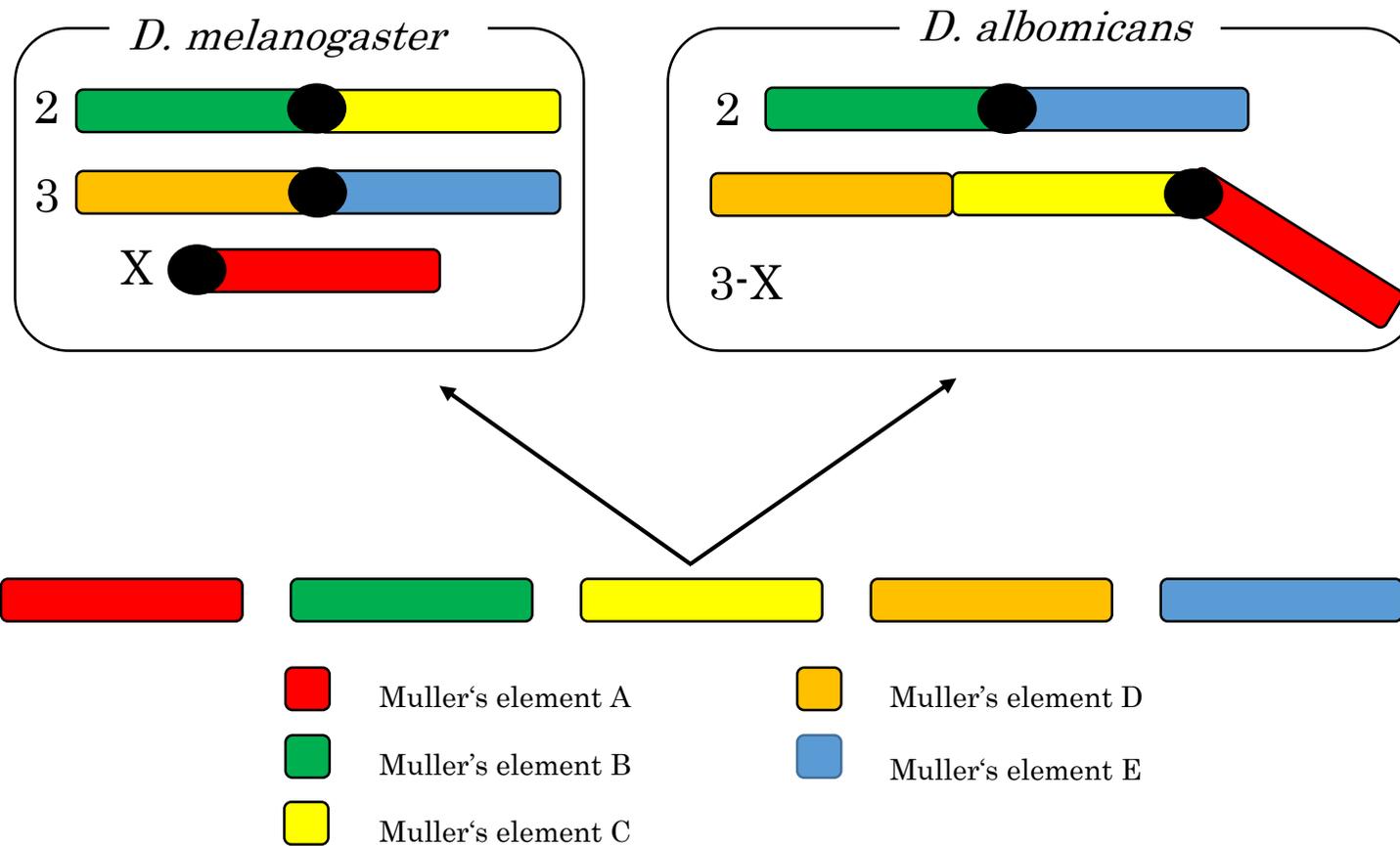


図3 キイロショウジョウバエおよびアカショウジョウバエの染色体構造の模式図

キイロショウジョウバエとアカショウジョウバエの染色体は異なる Muller's element が結合して形成されていることが知られている。アカショウジョウバエでは第3染色体と性染色体が融合した neo-sex 染色体が存在する。(Chang et al. 2008 に基づく)

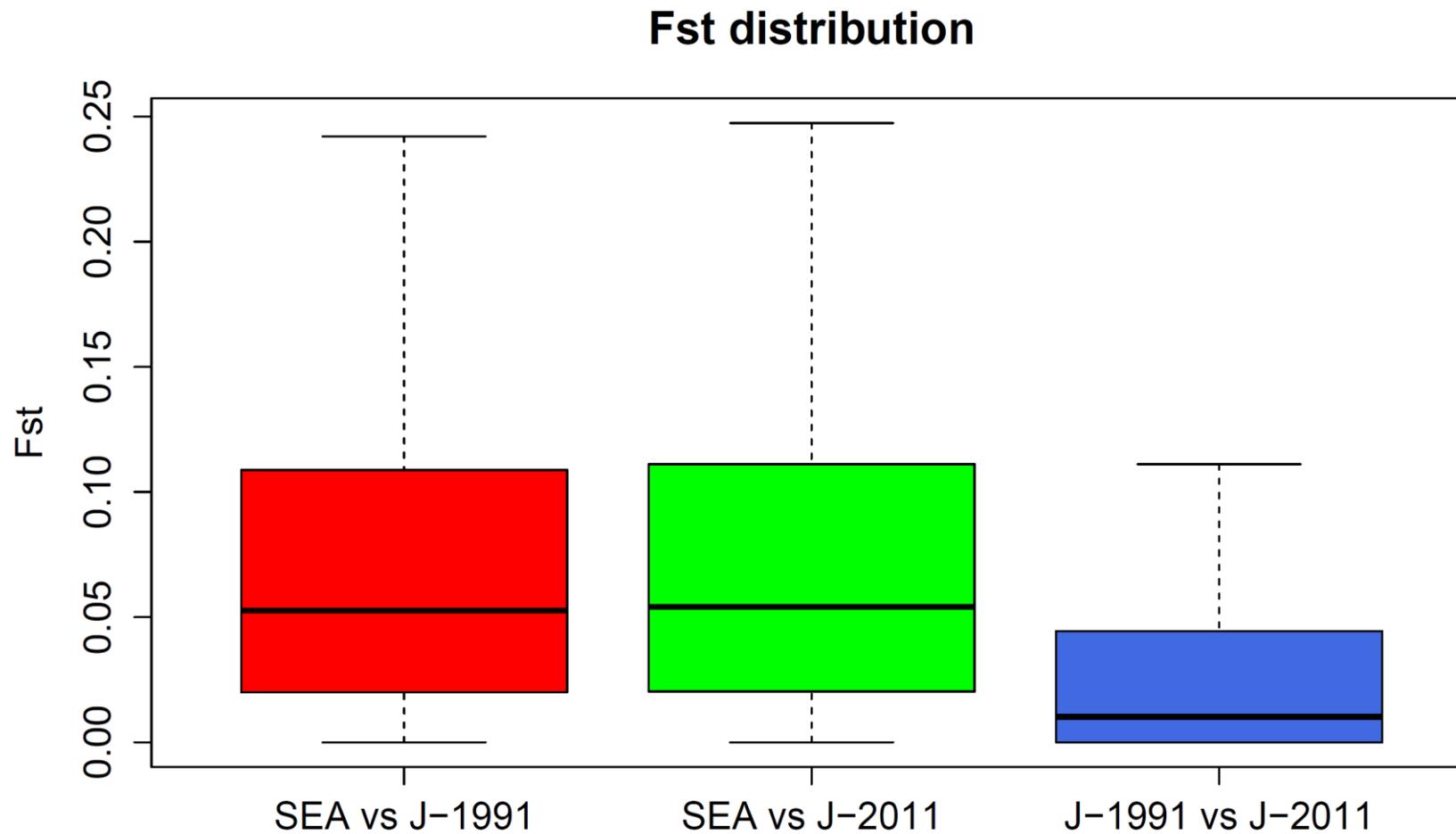


図4 各集団における SNP の  $F_{st}$  分布

SNP 数が多く、外れ値も多いため外れ値は除外している。SEA：東南アジア集団、J-1991：1991年西日本集団、J-2011：2011年西日本集団を示す。SNP サイトの総数は 276,421 だった。

## Tajima's D distribution

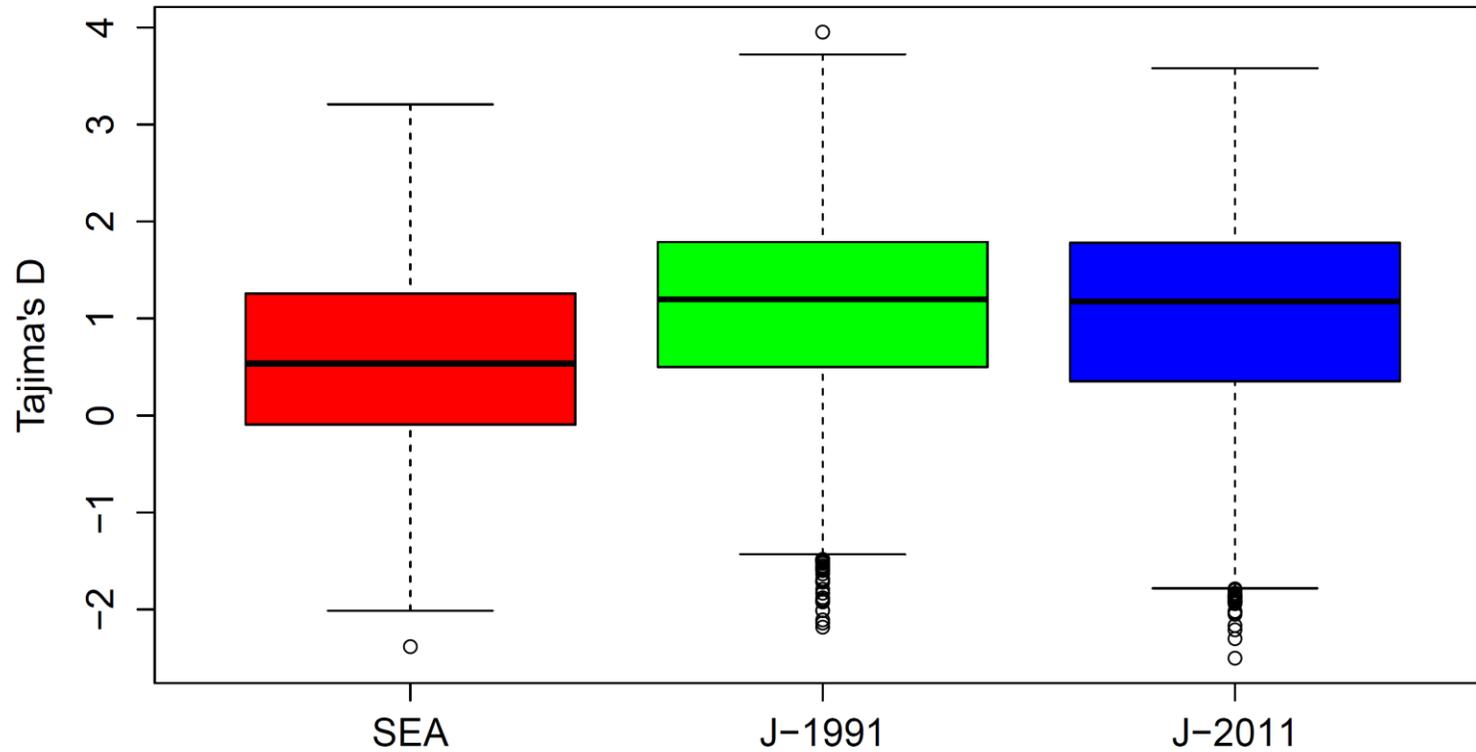


図5 8,540 遺伝子における Tajima's  $D$ 'の分布

GATK 内の haplotypcaller によって出力された VCF ファイルから R を用いて計算した。解析には遺伝子コドンの第 3 座位のみを用いた。

SEA : 東南アジア集団、J-1991 : 1991 年西日本集団、J-2011 : 2011 年西日本集団を示す。

## SEA vs Japan

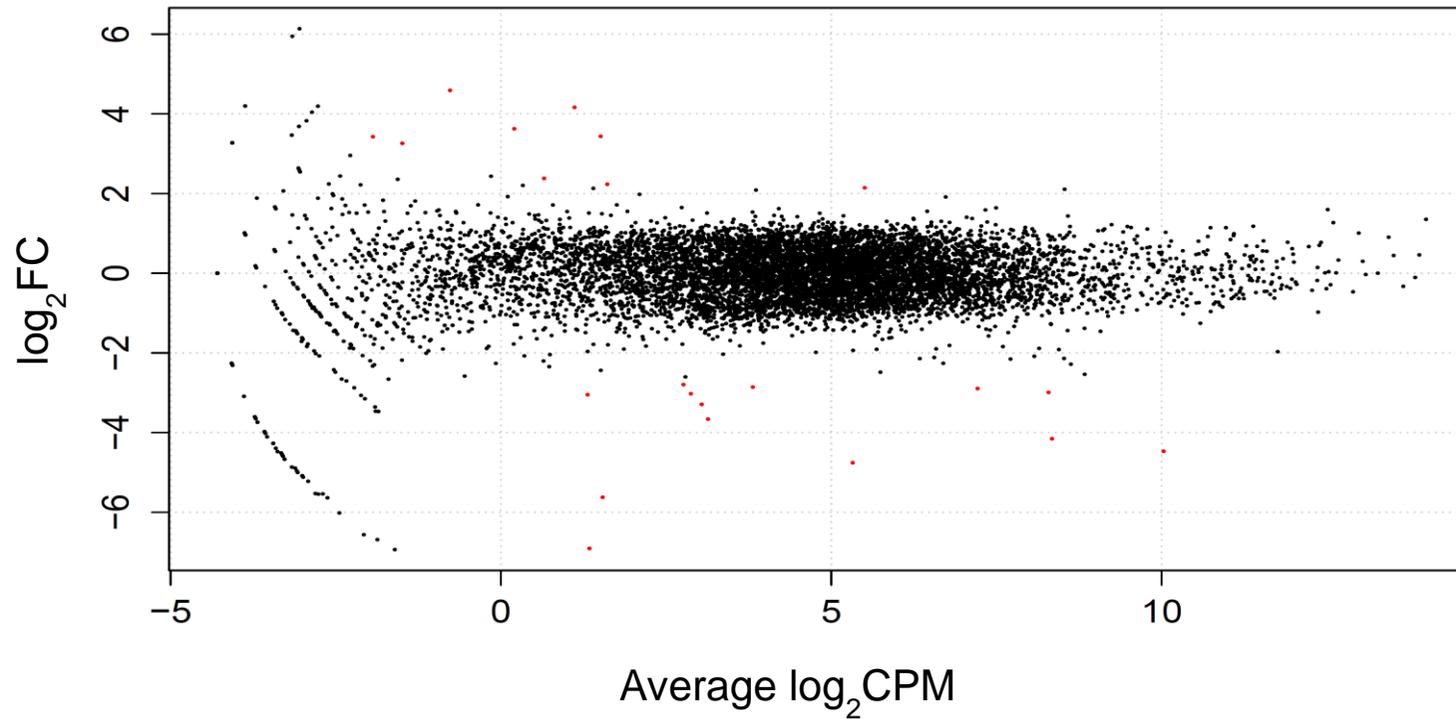


図6 東南アジア集団と西日本集団の8,540遺伝子の発現量のMAプロット

FCはFold change, CPMはCount pre millionを示す。また赤い点はFDR < 0.1以下の22遺伝子を示している。SEA:東南アジア集団 Japan:西日本集団2つを示す。

。

## Pi distribution

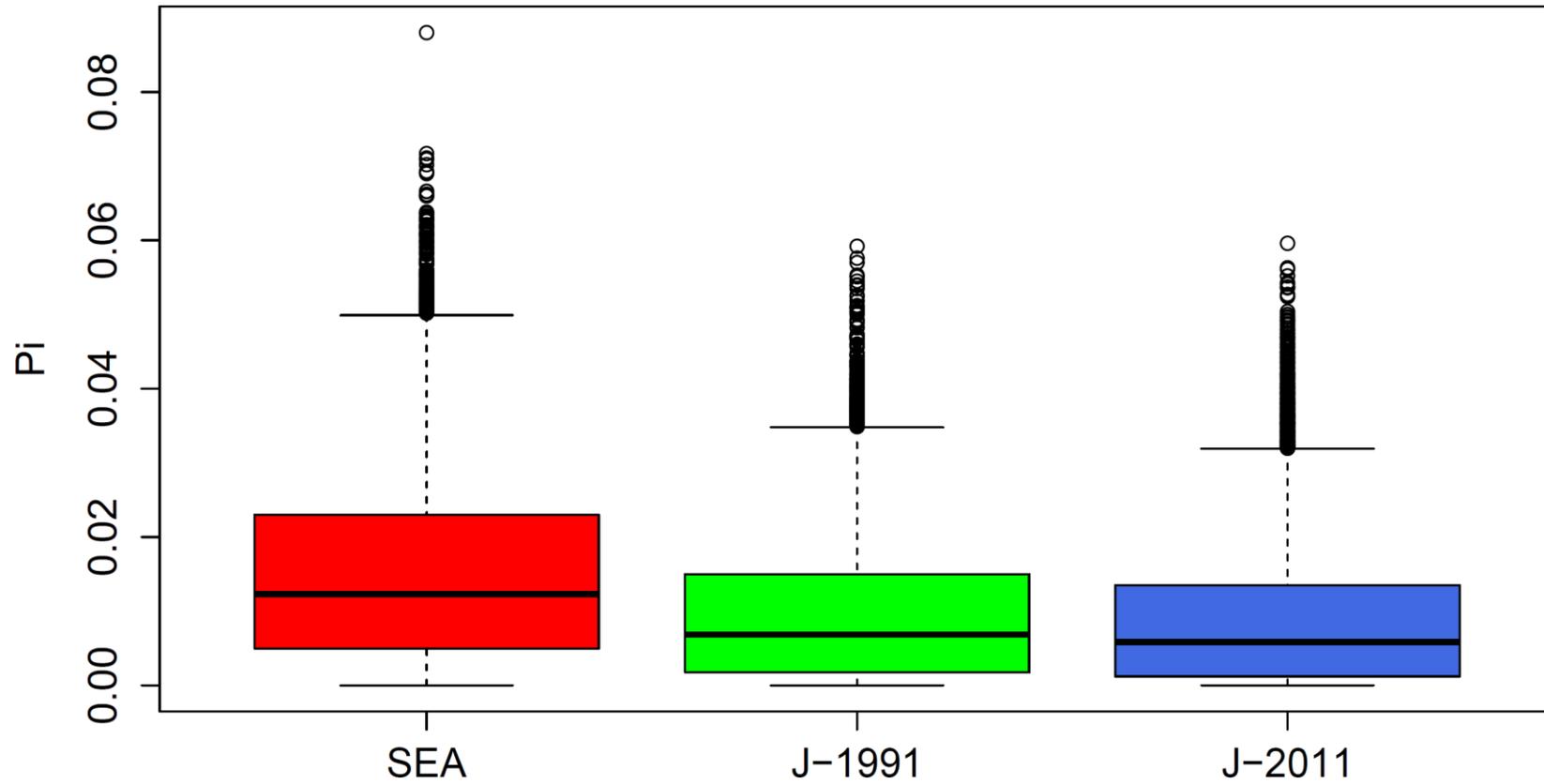


図 7a 8,540 遺伝子における  $\pi$  の分布

解析にはコドンの第3 座位のみを使っている。SEA : 東南アジア集団、J-1991 : 1991 年西日本集団、J-2011 : 2011 年西日本集団

## theta distribution

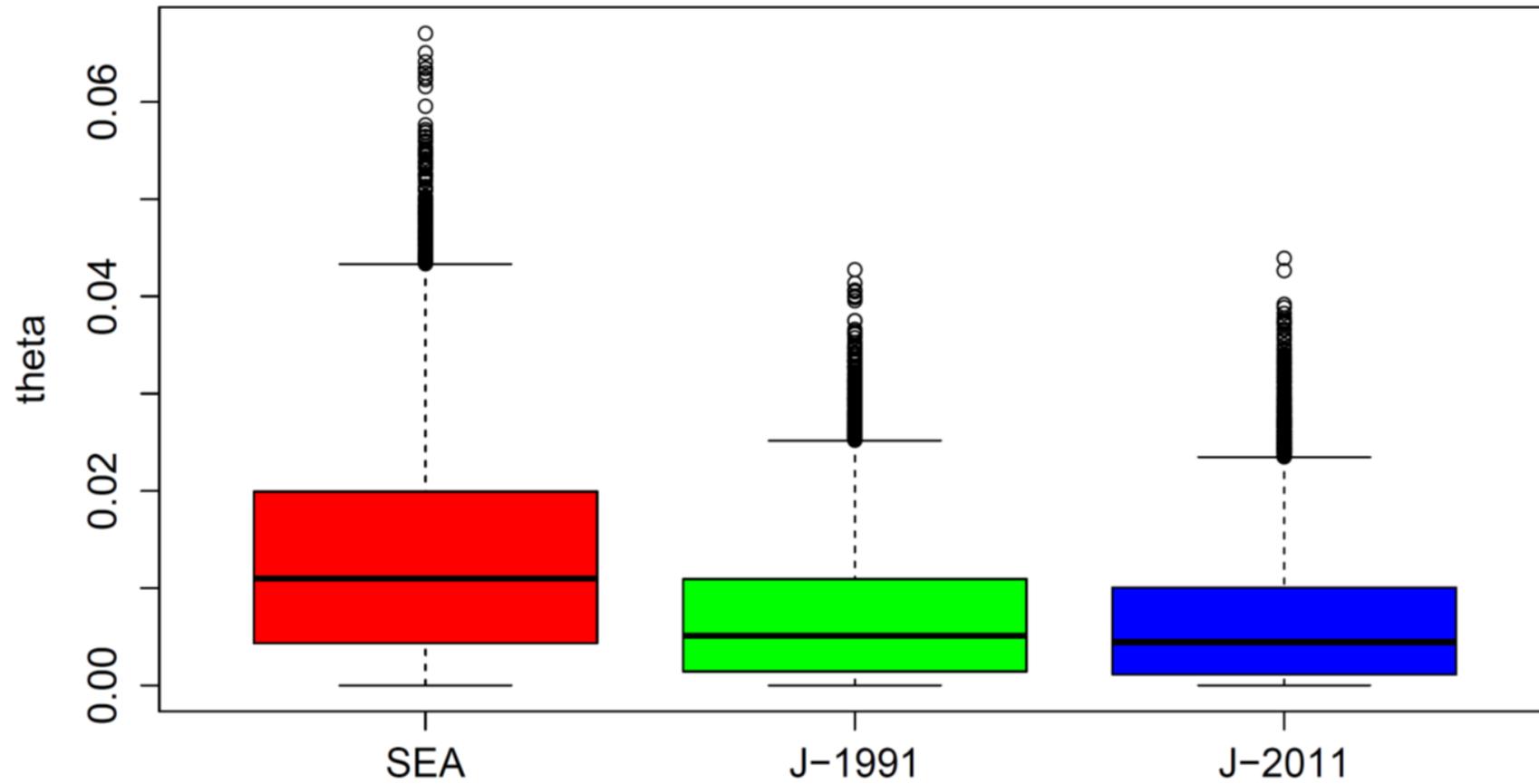


図 7b 8,540 遺伝子における  $\theta$  の分布

解析にはコドンの第3座位のみを使っている。SEA：東南アジア集団、J-1991：1991年西日本集団、J-2011：2011年西日本集団

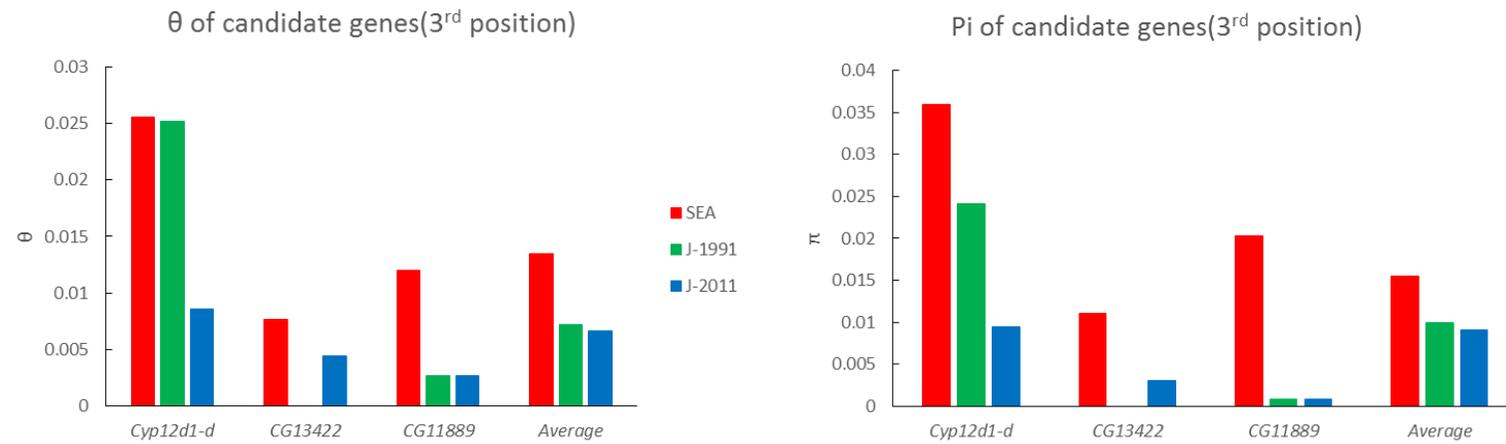


図8 低温耐性関連候補遺伝子および8,540遺伝子平均の $\theta$ および $\pi$

GATK内のhaplotypcallerで出力されたVCFファイルを基に計算した。解析には遺伝子コドンの第3座位のみを使用した。

SEA：東南アジア集団、J-1991：1991年西日本集団、J-2011：2011年西日本集団、Average：全遺伝子の平均

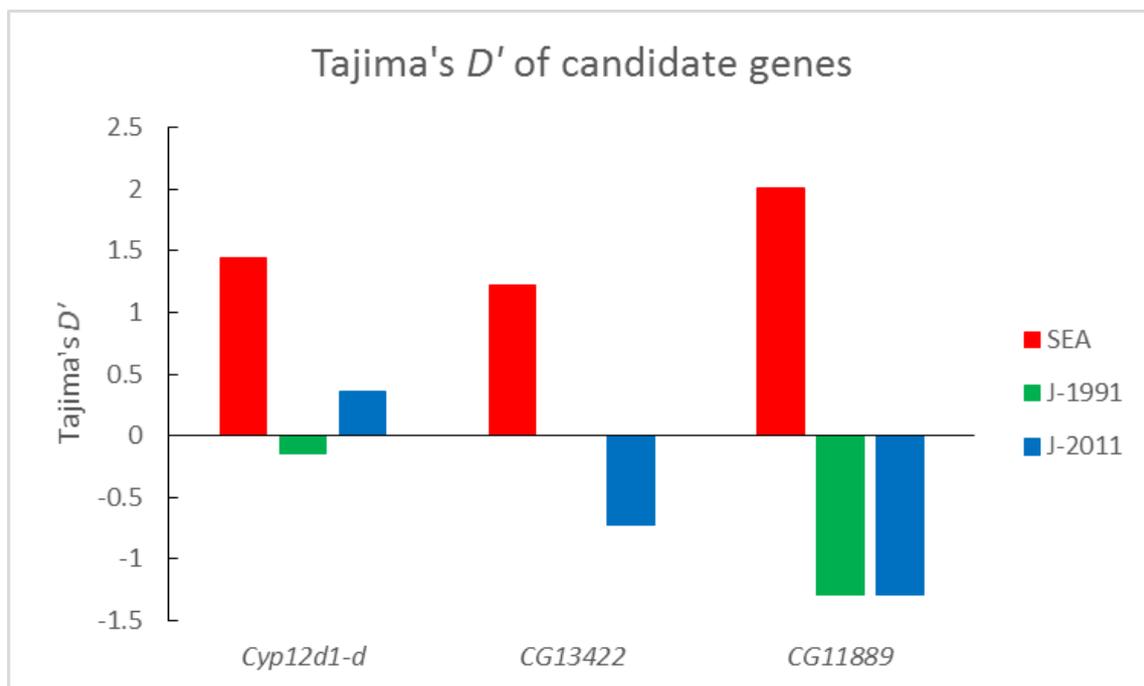


図9 低温耐性関連候補遺伝子における Tajima's  $D'$  の値

GATK 内の haplotypcaller で出力された VCF ファイルを基に計算した。解析には遺伝子コドンの第3座位のみを使用した。

SEA : 東南アジア集団、J-1991 : 1991年西日本集団、J-2011 : 2011年西日本集団を示している。

付表 1a pooled RNA-Seq に使用したアカシヨウジヨウバエ系統の採集地 (東南アジア集団)

		採集場所		採集年	系統名		
東南アジア	タイ	北部	チアンマイ	1977	CNX2		
					CNX29		
					CNX33		
				1982	CM5		
					CM130		
	マレーシア	西マレーシア	ペナン島	1979	Y53		
					Y57		
					Y67		
					クアラルンプール	1979	X3
						X13	
ミャンマー	北西	シュウェボー	1982	X86			
				SWB101			
				SWB102			
	中央	メイミョー	1982	SWB103			
				MMY1			
MMY2							
MMY3							

付表 1b pooled RNA-Seq に使用したアカシヨウジョウバエ系統の採集地 (1991 年西日本集団)

		採集場所		採集年	系統名		
東アジア	日本	近畿	紀伊勝浦	1991	KKU202		
					KKU204		
					KKU207		
					KKU210		
				兵庫	1991	SUM531	
			SUM601				
			SUM612				
			SUM616				
				中国	広島	1991	HRS402
			HRS421				
			HRS423				
			HRS435				
				四国	高松	1990	TAK401
			TAK412				
			TAK416				
			TAK418				
				九州	熊本	1990	KMT21
			KMT402				
			KMT412				
	KMT419						

付表 1c pooled RNA-Seq に使用したアカシヨウジョウバエ系統の採集地 (2011 年西日本集団)

		採集場所		採集年	系統名
東アジア	日本	中部	名古屋	2011	NGO2011-1
					NGO2011-2
					NGO2011-3
					NGO2011-4
					NGO2011-6
		近畿	京都	2011	ITM2011-1
					ITM2011-2
					ITM2011-3
					ITM2011-10
					ITM2011-11
		四国	兵庫	2011	UKB2011-1
					松山
			MYJ2011-2		
			MYJ2011-3		
			徳島	2011	TKS2011-1
TKS2011-2					
			TKS2011-4		
			TKS2011-5		

付表 2 pooled RNA-Seq 解析に用いた各プログラムのコマンド

プログラム	コマンド
Trinity	<code>--seqType fa --JM 20G --single &lt;input fas.file&gt; --output &lt;mame.out&gt; --CPU6</code>
Bowtie2	<code>-x &lt;index_file&gt; -f &lt;input_file&gt; -N 1 -S &lt;output.sam&gt;</code>
SAMtools	<code>view -q 20 -bS output.sam &gt; &lt;output_q20.bam&gt;</code>  <code>sort &lt;output_q20.bam&gt; &lt;output_q20_sort&gt;</code>  <code>mpileup -B &lt; output_q20_sort_remove &gt; &lt; output_q20_sort_remove &gt; &gt; &lt;output.mpilep&gt;</code>
Picard	<code>java -jar MarkDuplicates.jar ASSUME_SORTED=true REMOVE_DUPLICATES=true INPUT= output_q20_sort.bam OUTPUT=&lt;output_q20_sort_remove&gt; METRICS_FILE=duplicate</code>
PoPoolation2	<code>java -ea -Xmx150g --jar mpileup2sync.jar --input &lt;output.mpilep&gt; --output &lt;output.sync&gt; --fastq-type illumina --min-qual 5 --threads 12</code>  <code>perl snp-frequency-diff.pl --input &lt;output.sync&gt; --output-prefix &lt;output&gt; --min-count 4 -- min-coverage 20 --max-coverage 500</code>  <code>perl fst-sliding.pl --input &lt;output.sync&gt; --output &lt;output.fst&gt; --suppress-noninformative -- min-count 4 --min-coverage 20 --max-coverage 500 --min-coverd-fraction 1 --window-size 1 --pool-size 40</code>
GATK	<code>java -Xmx150G -jar GenomeAnalysisTK.jar -T HaplotypeCaller -R &lt;reference.fa&gt; -I &lt;output_q20_sort_remove.bam&gt; -ploidy 40 -stand_call_conf 10 -stand_emit_conf 30 -ncf 12 -o &lt;output.vcf&gt;</code>

全てのプログラムは Ubuntu ver. 12.04 上で使用した。

付表 3a アカシヨウジヨウバエ(オス)において低温順化の有無で大きく転写量が増加した遺伝子

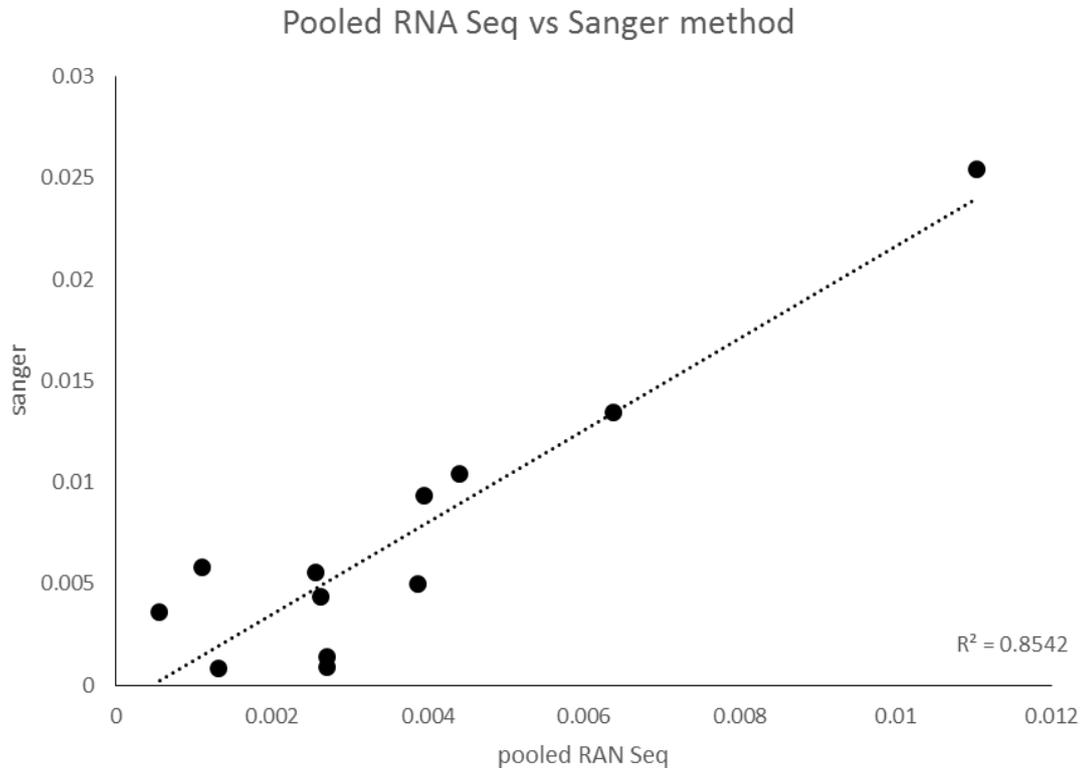
遺伝子名	RPKM		変化率
	低温順化なし(reads)	低温順化あり(reads)	
<i>Lip3</i>	1	13	8.51
<i>Jon25Bi</i>	11	80	7.24
<i>CG17374</i>	3	17	6.46
<i>Jon25Biii</i>	10	65	6.28
<i>Jhe</i>	4	24	6.18
<i>CG8952</i>	19	116	6.08
<i>Pepck</i>	95	562	5.93
<i>Jon99Cii</i>	12	71	5.76
<i>CG11796</i>	80	427	5.31
<i>Jon66Cii</i>	26	135	5.11
<i>CG16758</i>	28	145	5.10
<i>CG11796</i>	33	165	5.01
<i>CG42269</i>	1	3	5.00
<i>CG6295</i>	35	173	4.94
<i>CG5150</i>	5	24	4.70
<i>CG6271</i>	10	47	4.67
<i>CG12116</i>	11	49	4.43
<i>Jon44E</i>	4	19	4.39
<i>Acp1</i>	3	12	4.02
<i>CG12374</i>	92	354	3.86
<i>Prx2540-2</i>	38	145	3.79
<i>CG7916</i>	67	248	3.70
<i>LysD</i>	10	34	3.44
<i>kni</i>	4	13	3.26
<i>CG4734</i>	5	16	3.18
<i>CG4847</i>	166	525	3.17
<i>Jon99Fi</i>	11	35	3.16
<i>Sirup</i>	25	79	3.14
<i>CG1544</i>	3	9	3.11
<i>CG8997</i>	144	437	3.03
<i>CG1887</i>	3	10	3.01
<i>CG6910</i>	89	262	2.93
<i>Sdr</i>	10	30	2.89
<i>Npc2d</i>	27	78	2.86
<i>CG4020-PA</i>	16	44	2.72
<i>ste24c-PA</i>	4	12	2.71
<i>Jon65Aiv-PA</i>	248	671	2.71
<i>Chl9-PA</i>	13	35	2.70
<i>CG16758-PF</i>	112	296	2.65
<i>Amyrel-PA</i>	20	53	2.64
<i>SpdS-PA</i>	16	41	2.64
<i>aay-PA</i>	70	184	2.64
<i>Jheh1-PA</i>	34	88	2.63
<i>CG6839-PA</i>	5	14	2.62
<i>Oat-PA</i>	30	77	2.60
<i>CG1544-PA</i>	5	13	2.57
<i>LysC-PA</i>	3	7	2.57
<i>Cyp12d1-p-PA</i>	30	77	2.55
<i>U3-55K-PA</i>	4	11	2.55
<i>Tk-PA</i>	24	61	2.54
<i>CG32483-PA</i>	8	21	2.50

磯部(2014)から引用。低温順化したアカシヨウジヨウバエと低温順化していない成虫個体とでRNA-Seqを行い、得られたリード数の比較から順化の有無で転写量が大きく増加(2.5倍以上)した遺伝子を示している。遺伝子名は、キイロシヨウジヨウバエの相同遺伝子から引用している。

付表 3b アカシヨウジヨウバエ(オス)において低温順化の有無で大きく転写量が減少した遺伝子

遺伝子名	RPKM		変化率
	低温順化なし(reads)	低温順化あり(reads)	
<i>alpha-Est2</i>	103	18	0.17
<i>CG14153</i>	23	5	0.22
<i>Lsp2</i>	11	3	0.23
<i>CG6426</i>	98	24	0.25
<i>CG11889</i>	66	17	0.26
<i>tim-PJ</i>	63	17	0.26
<i>Gld</i>	13	3	0.27
<i>CG14191</i>	290	80	0.28
<i>CG11878</i>	172	51	0.30
<i>Mtk</i>	24	7	0.30
<i>CG16762</i>	106	33	0.31
<i>CG5172</i>	42	13	0.32
<i>eloF-PA</i>	30	10	0.34
<i>CG10513-PB</i>	105	37	0.35
<i>tim-PK</i>	106	37	0.35
<i>prc-PB</i>	8	3	0.35
<i>CG30280-PB</i>	10	4	0.36
<i>tim-PE</i>	47	17	0.37
<i>CG30281-PA</i>	185	71	0.38
<i>CG30280-PC</i>	25	9	0.39
<i>CG14963-PA</i>	45	17	0.39
<i>CG13422-PA</i>	350	138	0.39
<i>MtnC-PA</i>	118	48	0.40

磯部(2014)から引用。低温順化したアカシヨウジヨウバエと低温順化していない成虫個体とでRNA-Seqを行い、得られたリード数の比較から順化の有無で転写量が大きく減少(0.5倍以下)した遺伝子を示している。遺伝子名は、キイロシヨウジヨウバエの相同遺伝子から引用している。



付図 1 pooled RNA Seq とサンガー法の比較

pooled RNA Seq とサンガー法による SNP 同定の比較。pooled RNA Seq は GATK で出力された VCF ファイルから同定された SNP を自作の java プログラムから計算した。Sanger 法は ABI3130sequencer によって出力された配列データから MEGA6 (Tamura et al. 2013) によって計算した。使用した系統は pooled RNA Seq:2011 年西日本集団、Sanger 法:2011 年に採集された 16 系統 (MYj2011-01,MYJ2011-04,MYJ2011-09,MYJ2011-10,TKS2011-03,TKS2011-06,ITM2011-02,ITM2011-04,ITM2011-10,ITM2011-12,ITM2011-13,ITM2011-06,NGO2011-04,NGO2011-06,NGO2011-07,NGO2011-08)。使用した遺伝子は 12 遺伝子 (*tps*, *Tpi*, *SdhA*, *Pglym78*, *Pdk*, *men-b*, *Mdh2*, *kdn*, *Idh*, *Gpo-1*, *Got2*, *Gapdh1*, *aralar*) である。

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*, 30: 2725-2729.