

平成 30 年 6 月 11 日現在

機関番号：17102

研究種目：基盤研究(B) (一般)

研究期間：2014～2017

課題番号：26291082

研究課題名(和文) アンブリコンシーケンス解析を用いた樹木集団進化に関する集団遺伝学的研究

研究課題名(英文) Population genetic studies of evolution of tree populations using amplicon sequencing method

研究代表者

館田 英典 (TACHIDA, Hidenori)

九州大学・理学研究院・教授

研究者番号：70216985

交付決定額(研究期間全体)：(直接経費) 12,200,000円

研究成果の概要(和文)：針葉樹スギ、ヌマスギで変種の分化の過程で起こった集団構造の変化を、各種約100個体で50から100程度の遺伝子領域の塩基配列を決定することにより推定した。変種間の分化はスギでは32万年から85万年前、ヌマスギでは650万年前に起こっており、その後も変種間の遺伝子流動が続いていることが明らかになった。また小笠原固有種のシマイスノキでも父島と母島の集団間でこれらの変種間と同様の遺伝的分化が進んでいることがわかった。

研究成果の概要(英文)：We inferred the past demographic history of two conifer species, *Cryptomeria japonica* and *Taxodium distichum*, using target amplicon sequencing of 50~100 genes in ca. 100 individuals. We found that four genetic clusters of *C. japonica* populations diversified 0.32-0.85 MYA and three clusters of *T. distichum* populations diversified 3.6-6.5 MYA, the estimates being much older than previously thought. We also investigated population structure of *Distylium racemosum*, an important component in Japanese broad-leaf evergreen forests, and its close relative, *D. lepidotum*, endemic to Ogasawara Islands, using amplicon sequencing. Although differentiation between Japanese populations of *D. racemosum* was weak, we found clear differentiation between *D. racemosum* and *D. lepidotum* and also differentiation between two populations of *D. lepidotum*. We found a few candidate genes for local adaptation in *T. distichum* and *D. racemosum*.

研究分野：集団遺伝学

キーワード：集団遺伝学 集団構造 自然淘汰 樹木集団 アンブリコンシーケンシング イスノキ スギ ヌマスギ

1. 研究開始当初の背景

これまでに集団の複数個体のサンプルから得られた塩基配列データをもとに、最尤法やベイズ法を使って、その生物の過去の集団構造や遺伝的パラメーターの推定、適応的淘汰の検出等が行われてきた。これらの研究はサンガー法によって得られたデータをもとにしており、信頼度の高い配列情報に基づいていたが、この方法では解析できる遺伝子座数を増やすことは難しく、単純なモデルを仮定した集団構造の推定や、少数の適応淘汰候補遺伝子の解析しか行うことが出来なかった。

次世代シーケンシング (NGS) 技術の発展はこの状況を一変させ、ヒト、ショウジョウバヤアラビドプシシス等のモデル生物では、複数個体で全ゲノム配列を決定する集団ゲノミクス研究が可能となった。これにより、詳細な過去の集団構造や過去に起こった適応進化について多くの知見が得られるようになってきた。しかし全ゲノム配列を多数個体でしかも高精度で決定することは、まだ費用の面から難しい。特に非モデル生物では基本情報となる参照ゲノム配列が存在せず、またヒト等のように大量の配列決定が出来ない。そこでゲノムのランダムな領域を増幅する RAD 法等を用いてある程度ゲノムの濃縮を行い、研究が進められてきた。しかし RAD 法では、濃縮される領域を予め決められない、全個体で同じ領域の配列が決定されないこと等により、サンガー法によって得られたような信頼度の高い配列アラインメントデータを得ることは難しい。このためこれらの解析では、塩基多様度や遺伝的分化等の単純な集団統計量と現在の集団構造の推定しか行われておらず、生物集団の進化の理解には不可欠な歴史的集団構造モデルのパラメーター推定や、多くの遺伝子座での適応淘汰の検出は行われていなかった。

2. 研究の目的

(1) 樹木種のアンプリコンシーケンシングによる集団遺伝学的解析

対象種として我々のこれまでの研究から塩基多様度や遺伝的分化の程度等が分かっており、近縁種もしくは変種間で形態などの分化が見られる次の3樹種を選んだ。

ヌマスギ (*Taxodium distichum*)

北アメリカの湿地に生育するヌマスギには、流れの有る川岸に見られる bald cypress と、流れの殆どない沼に見られる pond cypress の2変種が存在する。我々のこれまでの研究から、bald cypress ではフロリダとミシシッピ川流域の集団間に弱い遺伝的分化が見られ、また2変種間でも遺伝的分化が見ることがわかってきた。しかしこれらの分化がどのような集団の歴史を経て形作られたかについては分かっていなかった。また2012年に北アメリカは数十年來の干ばつに見舞われ、ヌマスギ集団でも多くの個体で枯損が見られた。個体間には枯損の程度の差が

見られたので、この種については乾燥耐性に寄与する遺伝子を同定するために、枯損の程度と乾燥耐性候補遺伝子についての関連解析も行う。

スギ (*Cryptomeria japonica*)

日本の固有種スギには、太平洋側に生育するオモテスギと日本海側に生育するウラスギの2変種が存在する。津村らの研究から日本の天然林ではそれぞれの変種中に更に遺伝的に分化した2グループが存在することが示唆されていた。しかしこれら4グループがどのように形成されたか等についてはよく分かっていなかった。

イスノキ (*Distylium racemosum*) とシマイスノキ (*D. lepidotum*)

日本の照葉樹林に生育する高木極相種イスノキの集団は国内では殆ど分化していないが、乾性低木林に生育し形態の異なる小笠原固有の近縁種シマイスノキとの間には遺伝的分化が見られることが予備的な研究で分かっていた。しかしこれまでの解析では調べた遺伝子座数が少なく、詳細な過去の集団構造モデルのパラメーター推定や、適応淘汰の検出はできていなかった。

この研究ではPCR法によりゲノム中の100程度の遺伝子領域 (~500塩基) を約100個体で増幅し、それらの領域の塩基配列をNGSにより決定する (アンプリコンシーケンシング)。これにより一個体一領域について平均100以上のリード数を確保して、精度の高い配列アラインメントデータを得る。このようにして得られたデータを利用して、これらの種の過去の集団構造を推定するとともに、適応淘汰の検出を行うことによって、これらの種や変種がどのように進化して来たのかを明らかにする。

(2) シーケンスデータの進化学的解析法に関する研究

NGSから得られる大量のデータを処理して、過去の集団構造の推定に耐えうる配列アラインメントデータを得る必要がある。このためのパイプラインを構築し最適化する。また過去の集団構造推定ではモデルに基づく推定法を用いるが、使われるモデルでは何らかの簡略化が行われる。この簡略化の影響を評価する理論的研究を行う。

3. 研究の方法

(1) 樹木種のアンプリコンシーケンシングによる集団遺伝学的解析

各樹種でのサンプルされた個体数と場所は次のとおりである。

ヌマスギ

次の12集団から各8個体、合計96個体の葉をサンプルした。

bald cypress : テキサス2集団、ミシシッピ川流域4集団、フロリダ3集団

pond cypress : ミシシッピ川流域1集団、フロリダ2集団

スギ

津村らの研究から明らかになった遺伝的に分化した4グループのそれぞれを代表する、屋久島(オモテ)、静岡(オモテ)、立山(ウラ)、青森(ウラ)の4集団からそれぞれ24個体、合計96個体の葉をサンプルした。

イスノキ

イスノキ: 鹿児島、宮崎、福岡、対馬、高知集団からそれぞれ12個体をサンプルした。シマイスノキ: 父島、母島集団からそれぞれ20個体、合計96個体の葉をサンプルした。

これらの個体の葉のサンプルからDNAを抽出した。ヌマスギとイスノキについては、これまでに得られているEST配列をもとにして、新たにそれぞれ47及び144遺伝子領域を増幅するプライマーを設計した。スギについてはすでに津村らが設計したことから、144遺伝子領域を増幅するプライマーを選び使用した。これらのプライマーを使って、まず各種1個体でサンガー法を使って各遺伝子領域の配列を決定した。これらの塩基配列は、参照配列として次に説明するNGSから得られたデータ(今後リードと呼ぶ)をマップ(各リードをそれぞれの遺伝子領域ごとに集める作業)する際に使用した。

次に各個体のDNAを使い、図1にあるような方法で2回PCRを行なって個体識別インデックス配列やNGS用のアダプターを付加し、遺伝子領域を増幅する。ヌマスギについてはこの反応を各個体、各遺伝子領域で行なったが、個体数、遺伝子領域数の多いスギとイスノキではフリーダムの社のAccess Array™ Systemを使用し、48個体・48遺伝子領域の反応を同時に行なった。

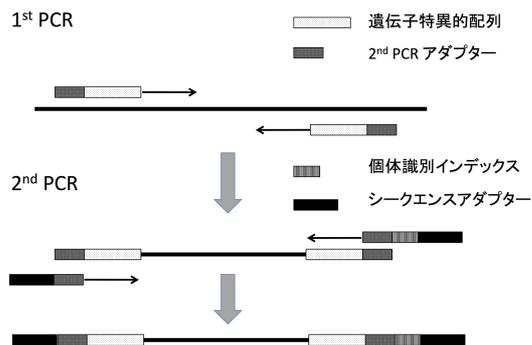


図1 PCRとタグ付加

増幅したDNAの濃度を調整し混合してNGSライブラリーを作成し、ヌマスギは454 GS-FLX、スギとイスノキはIllumina社のMiSeqを使用してNGSを行なった。

各種で個体識別インデックスを使って個体ごとにリードを分けたのち、ソフトウェアFASTX Toolkitを使ってQuality Scoreや長さに基づいて、得られたリードのquality controlを行なった。その後リードをソフトウェアBowtie 2またはBWAを使って参照配列にマップした。このデータからソフトウェアSAMtoolsを使い、各個体の各塩基サイトの遺伝子型を決定し、配列アラインメントを

得た。なおリード数が個体で10以下となるサイトの遺伝子型は欠損値とした。また各遺伝子領域で、各集団がHardy-Weinberg平衡(HWE)となっているかをどうか指標にして、パラログ遺伝子(重複遺伝子)を持つ遺伝子領域を除いた。なおNGS解析を行った中から少数個体、少数遺伝子座を選び、サンガー法による塩基配列決定を行って、NGSデータのエラー率を推定した。

得られた配列データを用いて、遺伝子型の主成分分析、ソフトウェアSTRUCTUREを使った集団構造解析を行い、各種でサンプルした個体が遺伝的にどのようにグループ化されるかを調べた後、塩基多様度、固定指数(F_{ST})、Tajima's D などの集団遺伝学的統計量を計算した。更にソフトウェアBayeScan等を使い、集団間の分化の程度から適応的変異を検出した。

Tajima's D やBayeScanで自然淘汰が検出されなかった遺伝子領域を中立進化遺伝子と考え、ヌマスギとスギについてはソフトウェアJaathaやfastsimcoal2を使って、過去の集団構造(集団の分岐年代、集団のサイズ、遺伝子流動の程度)を推定した。

2012年の北米での干ばつ後にテキサス集団からサンプルしたヌマスギ95個体からDNAを抽出し、乾燥耐性候補を含む141遺伝子領域で、上記のスギの場合と同じ方法でNGSにより塩基配列を決定した。これらの95個体については干ばつ後の枯損状況のデータもあるので、乾燥耐性遺伝子を見つける目的で、枯損の程度と遺伝子型との相関を調べた。

(2)シーケンシングデータの進化的解析法に関する研究

集団サイズが周期的に変動する場合を想定し、ソフトウェアmsを用いて遺伝子系図学シミュレーションを行った。そのデータをもとに、Tajima's D の平均などの統計量を使って変動の検出が可能かどうかを、変動周期、変動幅、調べる遺伝子座数、変異量を変えて調べた。

4. 研究成果

(1)樹木種のアンプリコンシーケンシングによる集団遺伝学的解析

ヌマスギ

96個体47遺伝子領域のNGSデータを得たが、HWE検定等によってパラログを持つ遺伝子領域を除いたため、最終的には39遺伝子領域の配列アラインメントを得ることができた。サンガー法による配列との比較を使ったerror率の推定値は0.04-0.13%だった。

このデータを使って現在の集団構造を解析したところ、ミシシッピー及びテキサスのbald cypress、フロリダのbald cypress、フロリダのpond cypressの3グループに分かれることがわかった。ミシシッピーの1pond cypress集団はbald cypress集団と同じグル

ープに属し形態情報からハイブリッド集団である可能性が疑われたので、以後の解析から除外した。BayeScan を使い適応変異の検出を行なったところ、bald cypress の地域間では3 遺伝子領域、変種間でも3 遺伝子領域で自然淘汰候補遺伝子領域が見つかった。

自然淘汰の候補遺伝子領域を除き、更に欠損データが少ない3 1 遺伝子領域のデータを用いて、Jaatha により過去の集団構造を推定した。その結果を図2 に示してある。bald cypress の地域集団分岐は約3 6 0 万年前に起こり、その後ミシシッピー・テキサス集団からフロリダへの遺伝子流動が続いている。またミシシッピー・テキサス集団はサイズの増大を続けている。一方変種の分岐は約6 5 0 万年前に起こり、分岐後も遺伝子流動が続いている。また pond cypress は分岐後サイズの拡大を続けている。これらの分岐は3 0 0 万年以上前で約1 0 万年の氷期サイクルよりずっと長い時間スケールで起こっており、分岐後の遺伝子流動によって地域間、変種間の中立遺伝子座での遺伝的分化が抑制されている。また分岐後に bald cypress のミシシッピー・テキサス集団と pond cypress で集団サイズの増大が起こっていることから、フロリダの bald cypress 集団が新しい生育地に進出することにより、新しい地域集団や変種が形成されたことが推測された。

テキサス集団9 5 個体で干ばつ後の枯損状況と、1 4 1 遺伝子領域の相関を調べたが、多重比較の補正を行なって有意となるSNP は見つからなかった。一方同時に調べた樹高と相関のあるSNP が1 個見つかった (FDR=0.1)。有意ではないが樹高と枯損状況には負の相関が見られるようなので、このSNP を含む遺伝子が乾燥耐性に関与している可能性がある。

スギ

9 6 個体1 4 3 遺伝子領域のNGS データを得たが、HWE 検定等によってパラログを持つ遺伝子領域が除かれたため、最終的には9 4 個体1 2 0 遺伝子領域のアラインメントデータを得ることができた。サンガー法による配列との比較を使った error 率の推定値は0.08-0.15%だった。

このデータを使って現在の集団構造を解析したところ、屋久島、静岡、立山+青森の3グループに分かれることがわかった。塩基配列データから推定した同義塩基多様度や F_{ST} の平均値は、過去に少数の核遺伝子座を使って推定された値とほぼ同じ値となった。Tajima's D が有意な値をとる遺伝子領域は見つからなかった。また BayeScan を使った適応変異の検出でも、適応淘汰候補遺伝子領域は見つからなかった。

そこで各集団からそれぞれ選んだ1 5 個体の1 2 0 遺伝子領域のデータを用い、fastsimcoal2 を使って4 集団の過去の集団構造の推定を行なった。その結果を図3 に示す。過去に現在の集団より大きな集団が存在

して約8 5 万年前に屋久島集団がそこから分岐し、更に約3 2 万年前に他の3 集団が分岐したことが推定された。尤度比検定により屋久島集団が他の集団より有意に前に分岐

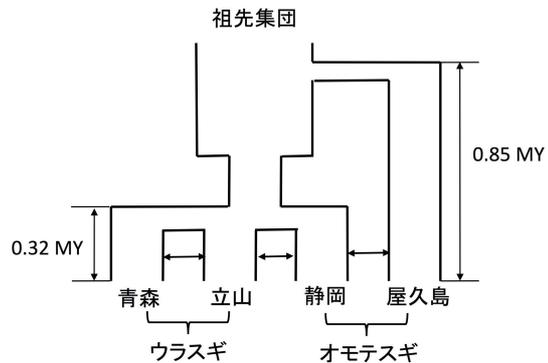


図3 スギの過去の集団構造。横向きの矢印は遺伝子流動を表している。

したことも示された。更に隣接集団間の遺伝子流動は現在も続いていることも推測された。屋久島集団が長い間分離されてきたことから、今回は検出することができなかったが、この集団には地域適応変異が蓄積されていると推測される。また屋久島集団を除く3 集団の分岐が、約1 0 万年前と考えられる直前の間氷期ではなく、その2 回前の間氷期に相当する時期に起こったと推定されたことも興味深い。

イスノキ

9 6 個体1 4 3 遺伝子領域のNGS データを得たが、HWE 検定等によりパラログを持つ遺伝子領域が除かれたため、最終的には9 5 個体の1 1 2 遺伝子領域のデータを得ることができた。サンガー法を使った配列との比較による error 率の推定値は0.025%だった。

このデータを使って現在の集団構造を主成分分析により調べた。その結果、日本のイスノキ集団間には遺伝的分化がほとんどないが、シマイスノキとイスノキの間には強い遺伝的分化があり、またシマイスノキの中でも父島と母島の集団間には遺伝的分化があることがわかった。実際に F_{ST} の平均値は、イスノキとシマイスノキ間で0.31、イスノキ集団間で0.01-0.05、シマイスノキの父島・母島集団間で0.14であった。

個々の遺伝子領域での F_{ST} 値を見ていくと、イスノキとシマイスノキの間で3 領域、父島と母島集団の間で1 領域で、 F_{ST} 値が0.7 を超えることがわかった。このような遺伝子領域は適応的な地域分化に寄与している可能性がある。他の2 樹種で行なった BayeScan を使った解析を行い、これらの F_{ST} 値が統計的に有意であるかどうかを確かめる必要がある。

ソフトウェア TreeMix を使い、マンサクを外群とした解析では、イスノキ集団からシマイスノキ集団がまず分岐し、その後母島と父島集団が分岐したと推定されたが、過去の集団構造を知るために、今後 fastsimcoal2 を

使った解析を行う必要がある。

(2) シークエンスデータの進化的解析法に関する研究

fastsimcoal2 などを使って過去の集団構造を推定する時には簡略化されたモデルを仮定するが、その際に繰り返し起こる集団サイズの変動など詳細な変化は取り入れられていない。そこで簡単な場合として、1 集団での周期的なサイズ変動がどのように遺伝的データに反映され、またその検出が可能かを調べた。その結果、 N 世代の間での変動回数が 0.5 から 100 までの間の変動は、100 遺伝子座程度のデータを得ると、Tajima's D の平均を使った方法でも fastsimcoal2 を使った方法でもかなりの確率で検出が可能だが、それ以外の変動回数では検出できないことがわかった。また集団サイズが小さい時点で得たデータを使うと、検出力が低下することもわかった。ちなみに Tajima's D の平均を使った方法と fastsimcoal2 を使った方法は検出力がほとんど変わらないが、計算時間は後者が 70 倍ほどになる。このことからまず Tajima's D の平均を使って詳細な変動を考慮するかどうかを判断した後にモデルを構築し、そのモデルを使って過去の集団構造を推定した方が効率的であることがわかった。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 5 件)

E. Moritsuka, P. Chhang, S. Tagane, H. Toyama, H. Sokh, T. Yahara, H. Tachida (2017) Genetic variation and population structure of a threatened timber tree *Dalbergia cochinchinensis* in Cambodia. *Tree Genetics & Genomes* 13: 115. DOI: 10.1007/s11295-017-1199-8 査読有

Y. Ikezaki, Y. Suyama, B. A. Middleton, Y. Tsumura, K. Teshima, H. Tachida, J. Kusumi (2016) Inferences of population structure and demographic history for *Taxodium distichum*, a coniferous tree in North America, based on amplicon sequencing analysis. *Am. J. Bot.* 103: 1937-1949. doi:10.3732/ajb.1600046 査読有

J. Kusumi, Y. Tsumura, H. Tachida (2015) Evolutionary rate variation in two conifer species, *Taxodium distichum* (L.) Rich var. *distichum* (baldcypress) and *Cryptomeria japonica* (Thunb. ex L.f.) D. Don (Sugi, Japanese cedar). *Genes & Genetic Systems* 90: 305-315. DOI: <http://doi.org/10.1266/ggs.14-00079> 査読有

M. Tamura, Y. Hisataka, E. Moritsuka, A. Watanabe, K. Uchiyama, N. Futamura,

K. Shinohara, Y. Tsumura, H. Tachida (2015) Analyses of random BAC clone sequences of Japanese cedar, *Cryptomeria japonica*. *Tree Genetics and Genomics* 11: 50. DOI 10.1007/s11295-015-0859-9 査読有

T. Yoshida, M. Tamekuni, T. Yahara, N. Inomata, H. Tachida (2014) Demographic history of a common pioneer tree, *Zanthoxylum ailanthoides*, reconstructed using isolation-with-migration model. *Tree Genetics and Genomics* 10:1213-1222. DOI 10.1007/s11295-014-0755-8 査読有

[学会発表](計 7 件)

森口 夏季、津村 義彦、手島 康介、楠見 淳子、館田 英典 他 4 名 アンブリコンシークエンス解析による塩基配列多型データに基づく針葉樹スギ (*Cryptomeria japonica*) の集団史の推定 遺伝学会第 89 回大会 2017 年

中村遥奈、手島康介、館田英典 塩基配列多型に基づく集団サイズの周期変動の検出 日本進化学会大 18 回東京大会 2016 年

池崎由佳、B. Middleton、津村義彦、手島康介、館田英典、楠見淳子 他 1 名 アンブリコンシークエンス解析による針葉樹ヌマスギの集団構造解析 日本進化学会第 18 回東京大会 2016 年

森口 夏季、津村 義彦、手島 康介、楠見 淳子、館田 英典 他 4 名 アンブリコンシークエンス解析による塩基配列多型データに基づく針葉樹スギ (*Cryptomeria japonica*) の集団構造の推定 遺伝学会第 88 回大会 2016 年

H. Tachida DNA polymorphisms and molecular evolution in trees. *Molecular Population Genetics and Evolution: Genes, Genomes and Models*. 2015

中村遥奈、手島康介、館田英典 集団サイズの周期変動による塩基配列多型への影響 遺伝学会第 87 回大会 2015 年

池崎由佳、B. Middleton、津村義彦、手島康介、館田英典、楠見淳子 他 1 名 アンブリコンシークエンス解析による針葉樹ヌマスギの集団構造、及び分化の歴史の推定 遺伝学会第 86 回大会 2014 年

[その他]

ホームページ等

6. 研究組織

(1) 研究代表者

館田 英典 (TACHIDA HIDENORI)

九州大学・理学研究院・教授

研究者番号：70216985

(2) 研究分担者

津村 義彦 (TSUMURA YOSHIHIKO)
(平成 28 年 8 月 25 日まで)
筑波大学・生命環境系・教授
研究者番号 : 20353774

手島 康介 (TESHIMA KOSUKE)
九州大学・理学研究院・助教
研究者番号 : 20447593

楠見 淳子 (20447593)
九州大学・比較社会文化研究院・准教授
研究者番号 : 20510522

(3) 連携研究者

()

研究者番号 :

(4) 研究協力者

ベス・ミドルトン (BETH MIDDLETON)
U.S. Geological Survey