

RAD-seq解析を中心とした target sequencing解析について



松波 雅俊

(琉球大学大学院医学研究科先進ゲノム検査医学講座)

自己紹介

- ・ 大阪府堺市出身
- ・ 北海道大学理学部生物学科卒業
- ・ 総研大遺伝学専攻博士課程修了：脊椎動物のゲノム重複の研究
- ・ 北海道大学でポスドク：有尾両生類の研究
- ・ 琉球大学大学院医学研究科：琉球列島人ゲノムの研究

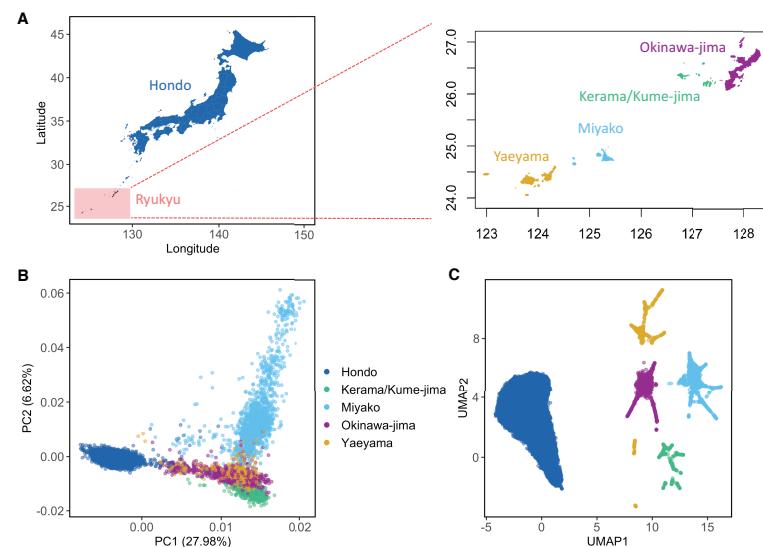


iNewt

integrated portal site for the Newt, *Pleurodeles waltl.*, database, genome editing, resource, and etcetera.

Natural Selection Signatures in the Hondo and Ryukyu Japanese Subpopulations

Xiaoxi Liu ^{1,2,†} Masatoshi Matsunami,^{3,‡} Momoko Horikoshi,⁴ Shuji Kunihiko Suzuki,⁵ Yukihide Momozawa ⁵, Shumpei Niida ⁶, Ryosuke Maeda ^{3,9}, Minako Imamura,^{3,9,*§} and Chikashi Terao ^{1,2,10,✉}



Most read

Beginner's Guide on the Use of PAML to Detect Positive Selection

Natural Selection Signatures in the Hondo and Ryukyu Japanese Subpopulations

Opsin Gene Duplication in Lepidoptera: Retrotransposition, Sex Linkage, and Gene Expression

Genomics Reveals Complex Population History and Unexpected Diversity of Eurasian Otters (*Lutra lutra*) in Britain Relative to Genetic Methods

Major Genetic Risk Factors for Dupuytren's Disease Are Inherited From Neandertals

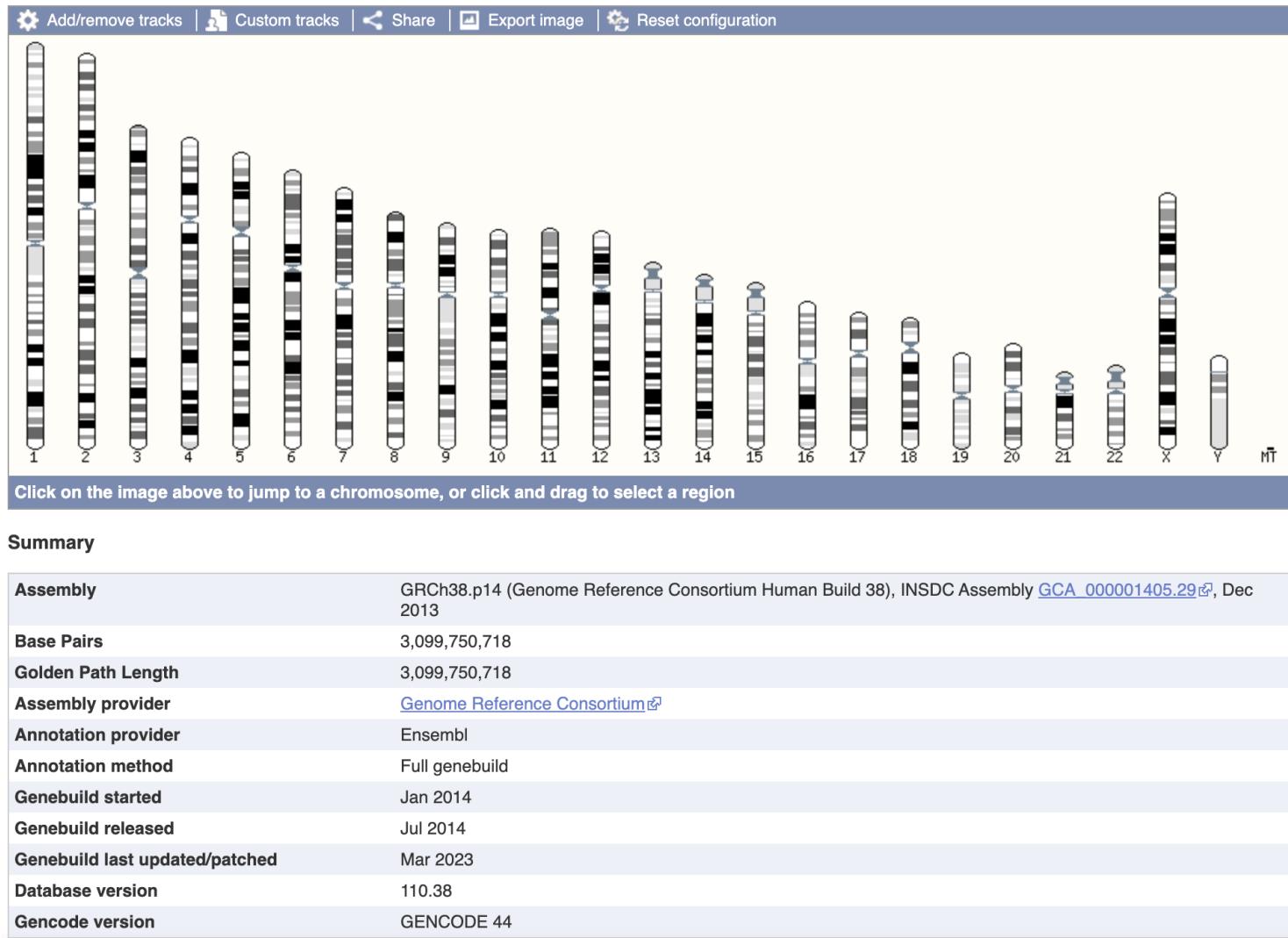
講義内容

- Target sequenceの原理と種類
- RAD-seq解析tips
- 実例紹介

Password
iBioArch_231210

Target sequenceの原理と種類

ゲノムは情報量が多い



https://asia.ensembl.org/Homo_sapiens/Location/Genome

何個体も全ゲノム解読してゲノム多様性を解析するのはコストの面からも難しいことがある
=> 領域を絞って解読する(target sequences)

遺伝多型(遺伝マーカー)に注目



多型(個人差)

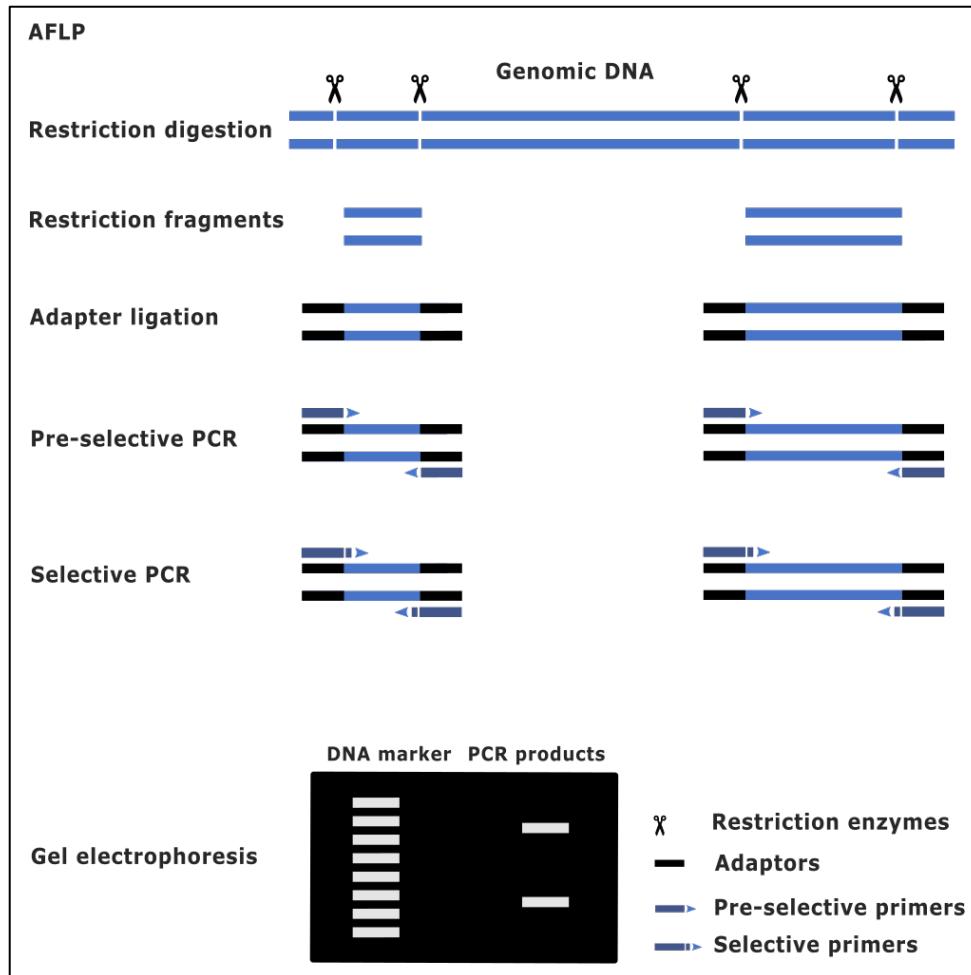
AGCTGATGGAT....TAACCGTATCC



AGCTGATAGAT....TAACCGTATCC

- ・ ゲノム上の塩基配列は個体間で少しずつ異なる。
- ・ ゲノム上の個人差を多型と呼ぶ。代表的なのは、一塩基単位で多型が生じる一塩基多型である。
- ・ 多型がある部分だけ効率的に読むことが出来ればコスパが良い。

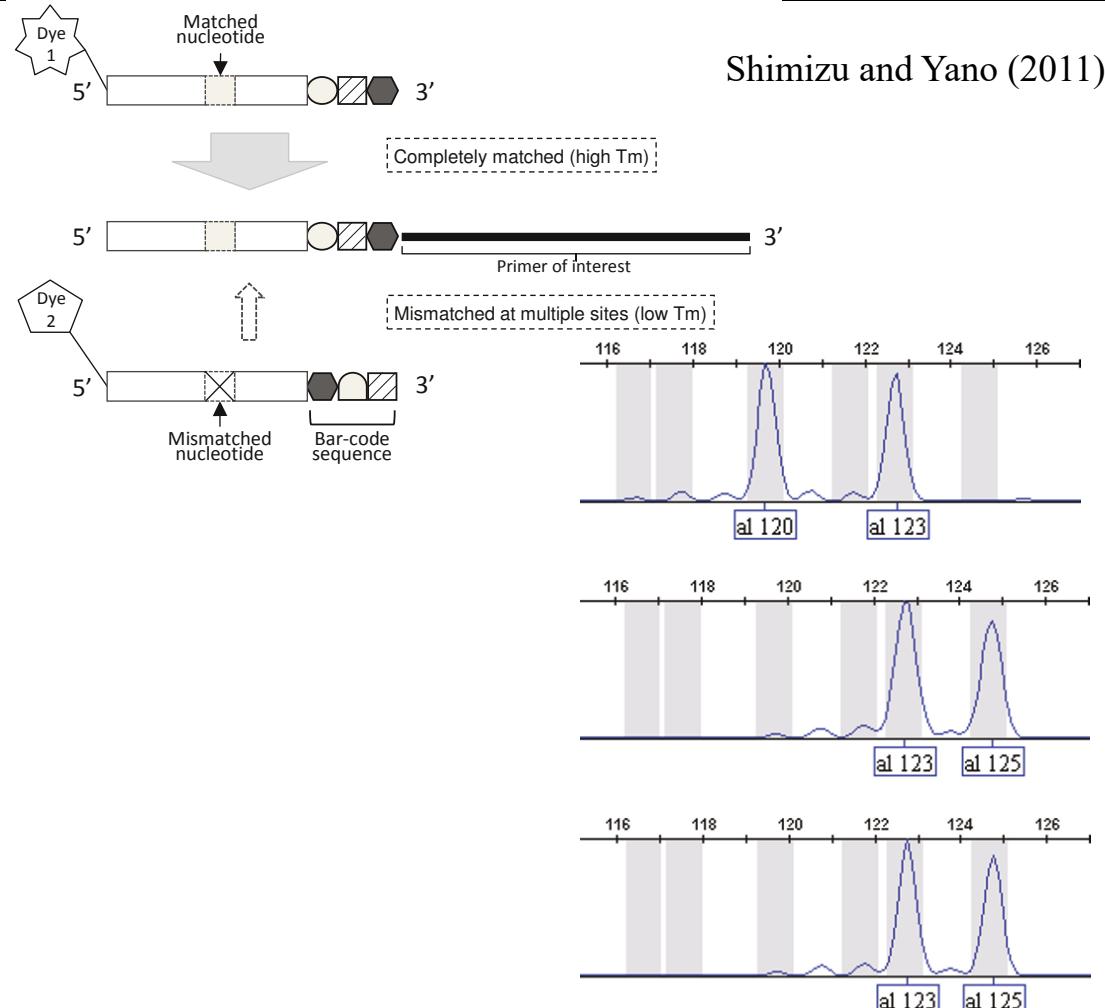
古典的な方法:AFLP (Amplified Fragment Length Polymorphism)



制限酵素により切斷したDNA断片をPCRにより増幅した際の增幅断片長のパターンによる多型、またその検出方法。

古典的な方法: microsatellite markers

5' -ATGC**ATGC**ATGC**ATGC**ATGC**ATGC**ATGC-3'
 5' -ATGC-----3'
 5' -ATGC**ATGC**ATGC**ATGC**ATGC**ATGC**-3'
 5' -ATGC**ATGC**ATGC-----3'
 5' -ATGC-----3'
 5' -ATGC**ATGC**ATGC**ATGC**ATGC**ATGC**-3'
 5' -ATGC**ATGC**ATGC-----3'
 5' -ATGC**ATGC**ATGC**ATGC**ATGC**ATGC**-3'
 5' -ATGC-----3'
 5' -ATGC**ATGC**ATGC**ATGC**ATGC**ATGC**-3'
 5' -ATGC-----3'
 5' -ATGC**ATGC**ATGC-----3'
 5' -ATGC-----3'
 5' -ATGC**ATGC**ATGC**ATGC**ATGC**ATGC**-3'
 5' -ATGC**ATGC**ATGC-----3'



Shimizu and Yano (2011)

ゲノム上の2~4個の塩基を一単位として複数回繰り返している部位をマイクロサテライト呼ぶ。このマイクロサテライトの繰り返し数は、同じ部位でも個体間によって異なるため、ゲノム多様性解析に用いられる。

利点: 低成本でゲノム多様性解析が可能。

欠点: Markerを見つけてprimerを作るのは結構大変。

古典的な方法: mitochondrial genome

Add/remove tracks | Custom tracks | Share | Export image | Reset configuration

Click on the image above to jump to a chromosome, or click and drag to select a region.

Summary

Assembly	GRCh38.p14 (Genome Reference Consortium 2013)
Base Pairs	3,099,750,718
Golden Path Length	3,099,750,718
Assembly provider	Genome Reference Consortium
Annotation provider	Ensembl
Annotation method	Full genebuild
Genebuild started	Jan 2014
Genebuild released	Jul 2014
Genebuild last updated/patched	Mar 2023
Database version	110.38
Gencode version	GENCODE 44

405.29 ↗, Dec

Pakendorf B, Stenzel M. 2005.
Annu. Rev. Genomics Hum. Genet. 6:165-83

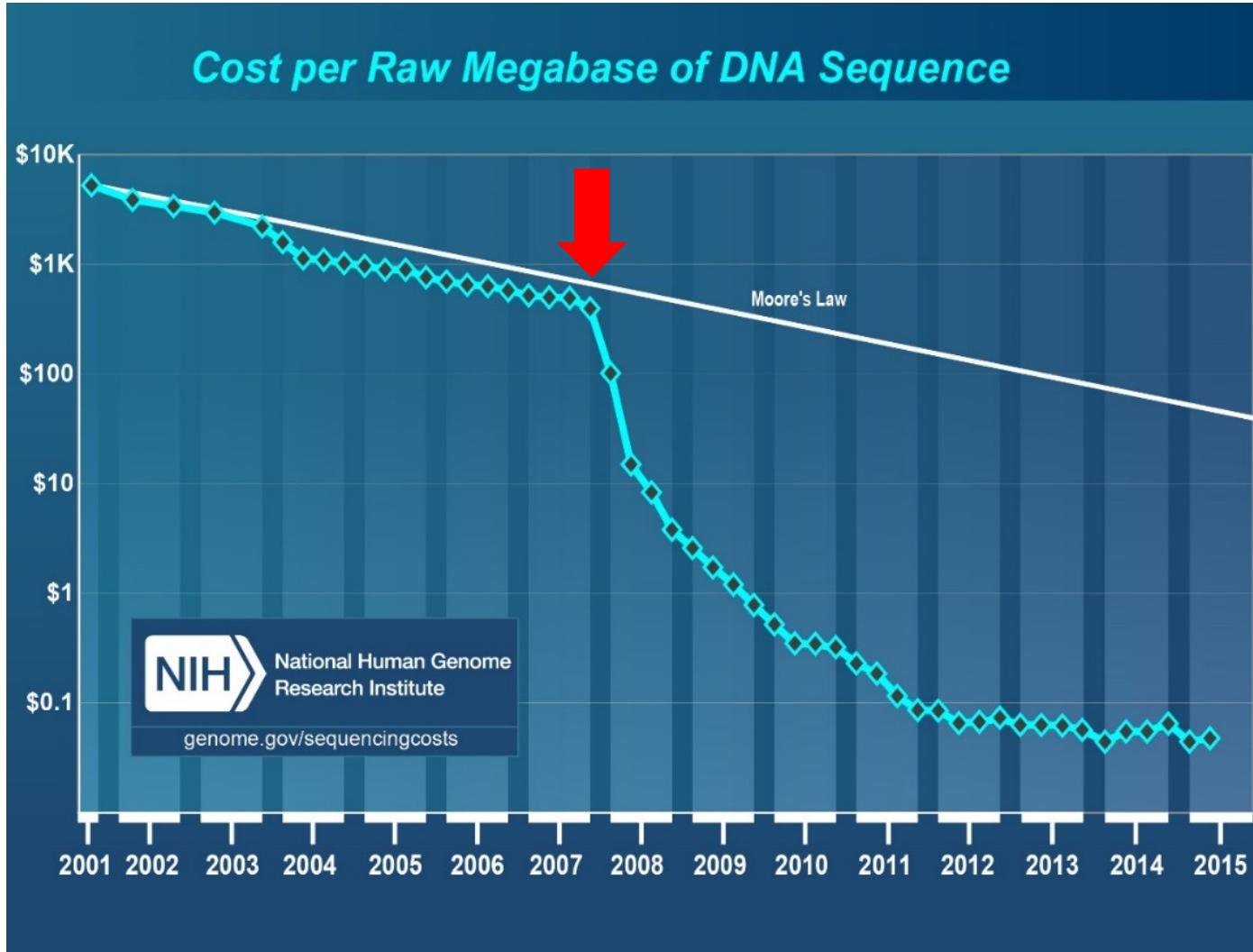
https://asia.ensembl.org/Homo_sapiens/Location/Genome

利点: 低成本で簡単に解読が可能。
欠点: 全ゲノムに比べて情報量が少ない!!!!

大規模シークエンサーの登場



シークエンス費用の変化



<http://www.genome.gov/sequencingcosts/>

ゲノム規模での遺伝多型の検出が可能な時代になった。

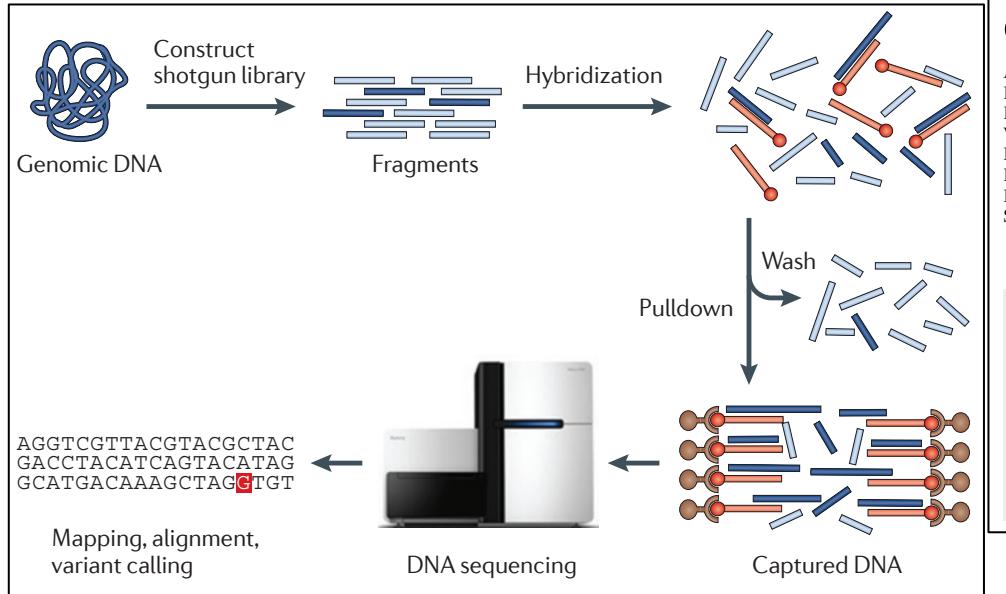
Short-read sequencerを用いたtarget seq方法

- Exome sequencing
- MIG sequencing (Multiplexed ISSR genotyping by sequencing)
- GRAS-Di (Genotyping by Random Amplicon Sequencing-Direct)
- Developed MAAS(multiple arbitrary amplicon sequencing)
- UCE (ultra-conserved element) target capture
- **RAD sequencing (Restriction site Association DNA sequencing)**

Exome sequencing

ARTICLE

<https://doi.org/10.1038/s41586-019-1457-z>



Bamshad et al. (2011)

Exome sequencing of Finnish isolates enhances rare-variant association power

Adam E. Locke^{1,2,3,43}, Karyn Meltz Steinberg^{2,4,43}, Charleston W. K. Chiang^{5,6,7,43}, Susan K. Service^{5,43}, Aki S. Havulinna^{8,9}, Laurel Stell¹⁰, Matti Pirinen^{8,11,12}, Haley J. Abel^{2,13}, Colby C. Chiang², Robert S. Fulton², Anne U. Jackson², Chul Joo Kang², Krishna L. Kanchi², Daniel C. Koboldt^{2,14,15}, David E. Larson^{2,13}, Joanne Nelson², Thomas J. Nicholas^{2,16}, Arto Pietilä⁹, Vasily Ramensky^{5,17}, Debashree Ray^{3,18}, Laura J. Scott³, Heather M. Stringham⁷, Jagadish Vangipurapu¹⁹, Ryan Welch³, Pranav Yajnik³, Xianyong Yin³, Johan G. Eriksson^{20,21,22}, Mika Ala-Korpela^{23,24,25,26,27,28}, Marjo-Riitta Järvelin^{29,30,31,32,33}, Minna Männikkö^{30,34}, Hannele Laivuori^{7,35,36}, FinnGen Project³⁷, Susan K. Dutcher^{2,13}, Nathan O. Stitzel^{2,38}, Richard K. Wilson^{2,14,15}, Ira M. Hall^{1,2}, Chiara Sabatti^{9,39}, Aarno Palotie^{7,40,41}, Veikko Salomaa⁹, Markku Laakso^{19,42}, Samuli Ripatti^{7,11,41}, Michael Boehnke^{3,44*} & Nelson B. Freimer^{5,44*}

Article

Exome sequencing and characterization of 49,960 individuals in the UK Biobank

<https://doi.org/10.1038/s41586-020-2853-0>

Received: 14 March 2019

Accepted: 25 August 2020

Published online: 21 October 2020

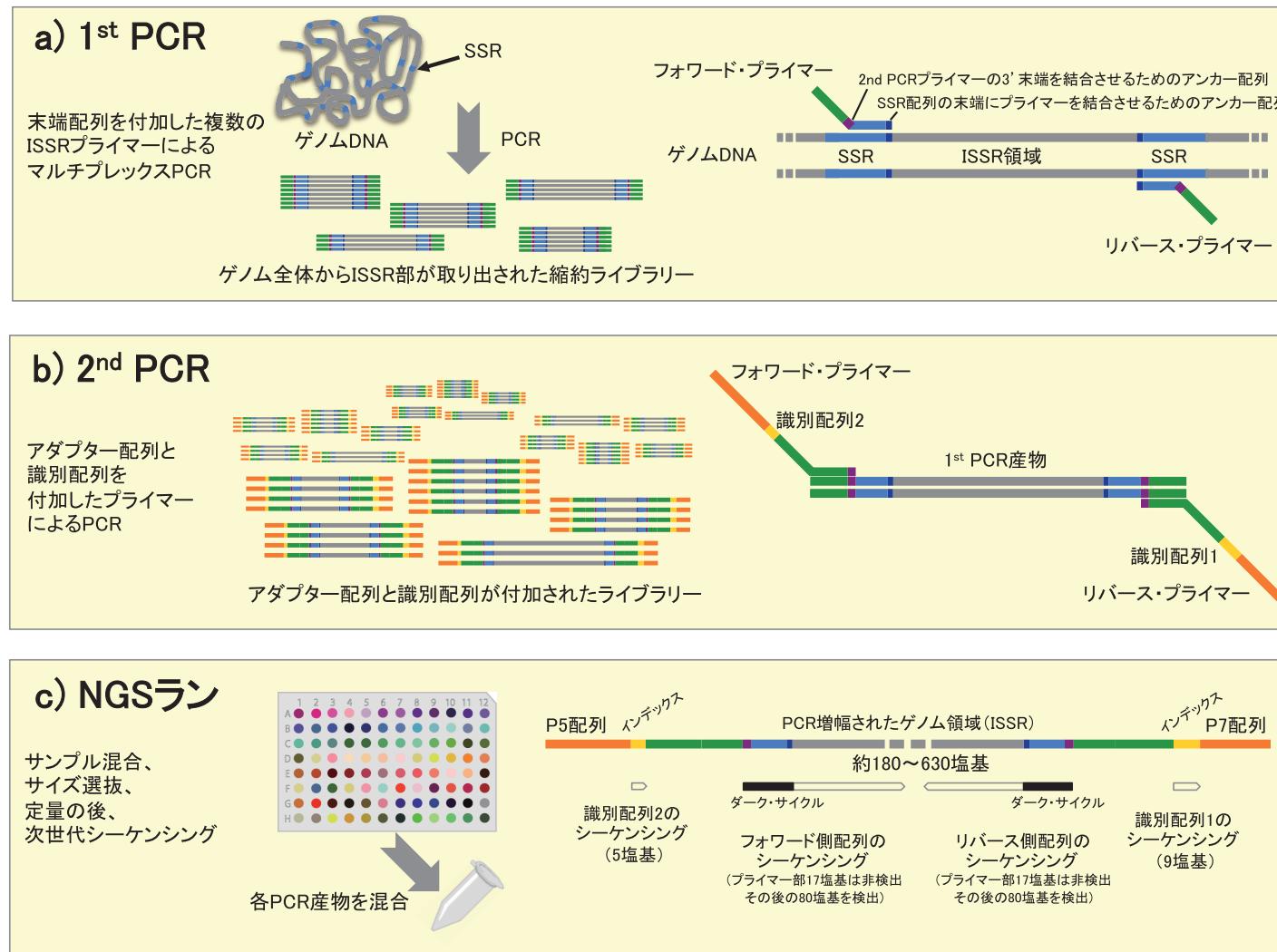
Open access

Check for updates

Cristopher V. Van Hout^{1,2}, Ioanna Tachmazidou^{2,10}, Joshua D. Backman¹, Joshua D. Hoffman^{3,11}, Daren Liu¹, Ashutosh K. Pandey³, Claudia Gonzaga-Jauregui¹, Shareef Khalid¹, Bin Ye¹, Nilanjana Banerjee¹, Alexander H. Li¹, Colm O'Dushlaine¹, Anthony Marcketta¹, Jeffrey Staples¹, Claudia Schurmann^{1,12,13}, Alicia Hawes¹, Evan Maxwell¹, Leland Barnard¹, Alexander Lopez¹, John Penn^{1,14}, Lukas Habegger¹, Andrew L. Blumenfeld¹, Xiaodong Bai¹, Sean O'Keeffe¹, Ashish Yadav¹, Kavita Praveen¹, Marcus Jones⁴, William J. Salerno¹, Wendy K. Chung^{5,6}, Ida Surakka⁷, Cristen J. Willer⁷, Kristian Hveem⁸, Joseph B. Leader⁹, David J. Carey⁹, David H. Ledbetter⁹, Geisinger-Regeneron DiscovEHR Collaboration*, Lon Cardon², George D. Yancopoulos⁴, Aris Economides⁴, Giovanni Coppola¹, Alan R. Shuldiner¹, Suganthi Balasubramanian¹, Michael Cantor¹, Regeneron Genetics Center*, Matthew R. Nelson^{3,15,16}, John Whittaker^{2,16}, Jeffrey G. Reid^{1,16}, Jonathan Marchini^{1,16}, John D. Overton^{1,16}, Robert A. Scott^{2,16}, Gonçalo R. Abecasis^{1,16}, Laura Yerges-Armstrong^{3,16} & Aris Baras^{1,16}

ヒトゲノム多様性解析では、WGSよりは低価なexome sequencingをバイオバンクスケール(数万人規模)で実施し、頻度の低いrare-variantまで調べるのが一般的になりつつある。

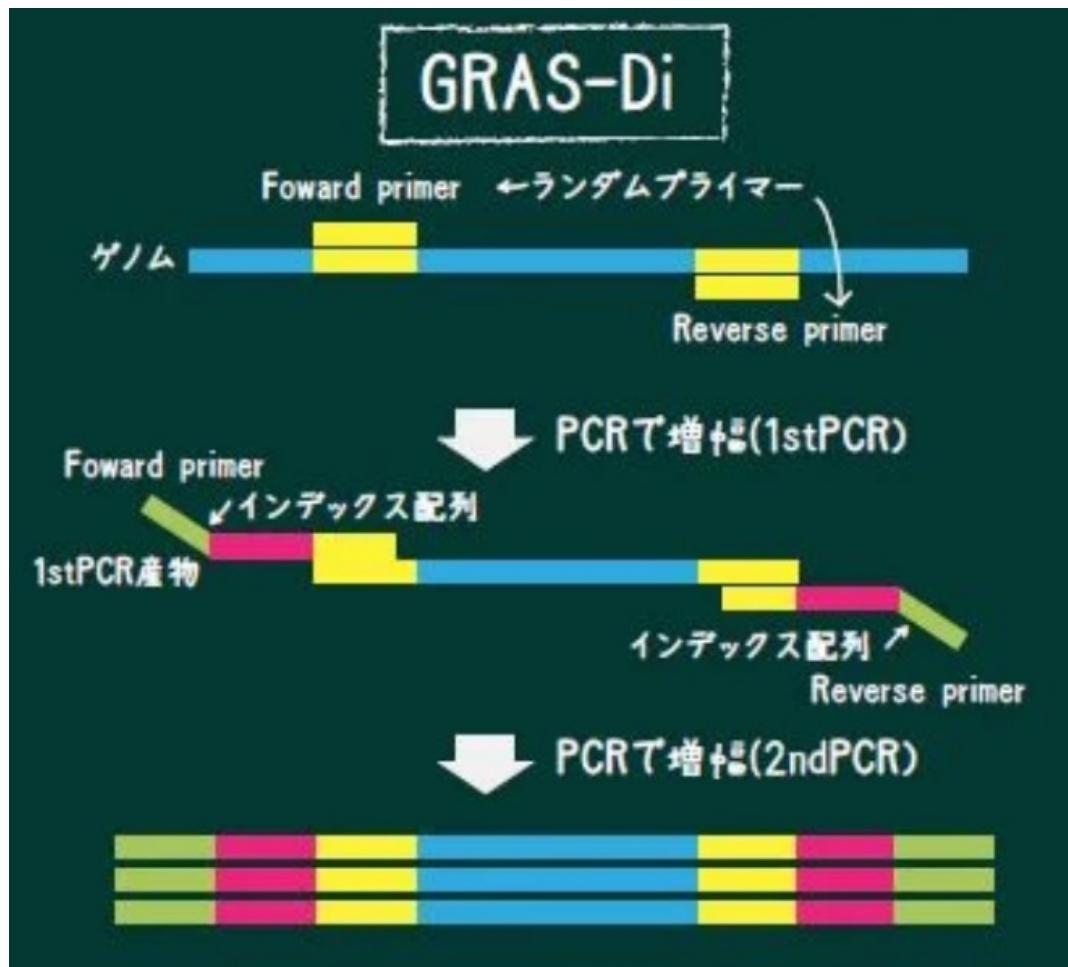
MIG sequencing (Multiplexed ISSR genotyping by sequencing)



陶山 (2019)

マイクロサテライト配列のような単純反復配列に挟まれた領域 (ISSR: inter-simple sequence repeat)を利用して SNP を検出する方法。PCR ベースなので微量・低品質な DNA 量で配列情報を取得することが可能である一方で、得られる SNP の数は少なく、PCR によるデータのバイアスが生じる危険性がある。
高品質な DNA を扱うことが難しいときに重宝する手法。

GRAS-Di (Genotyping by Random Amplicon Sequencing-Direct)

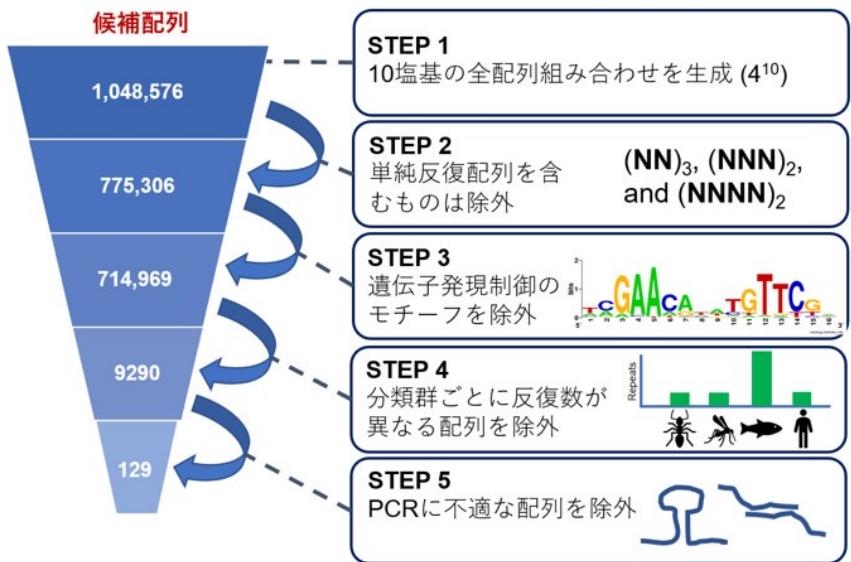


<https://gikenbio.com/dnaanalysis/ngs/radseq/>

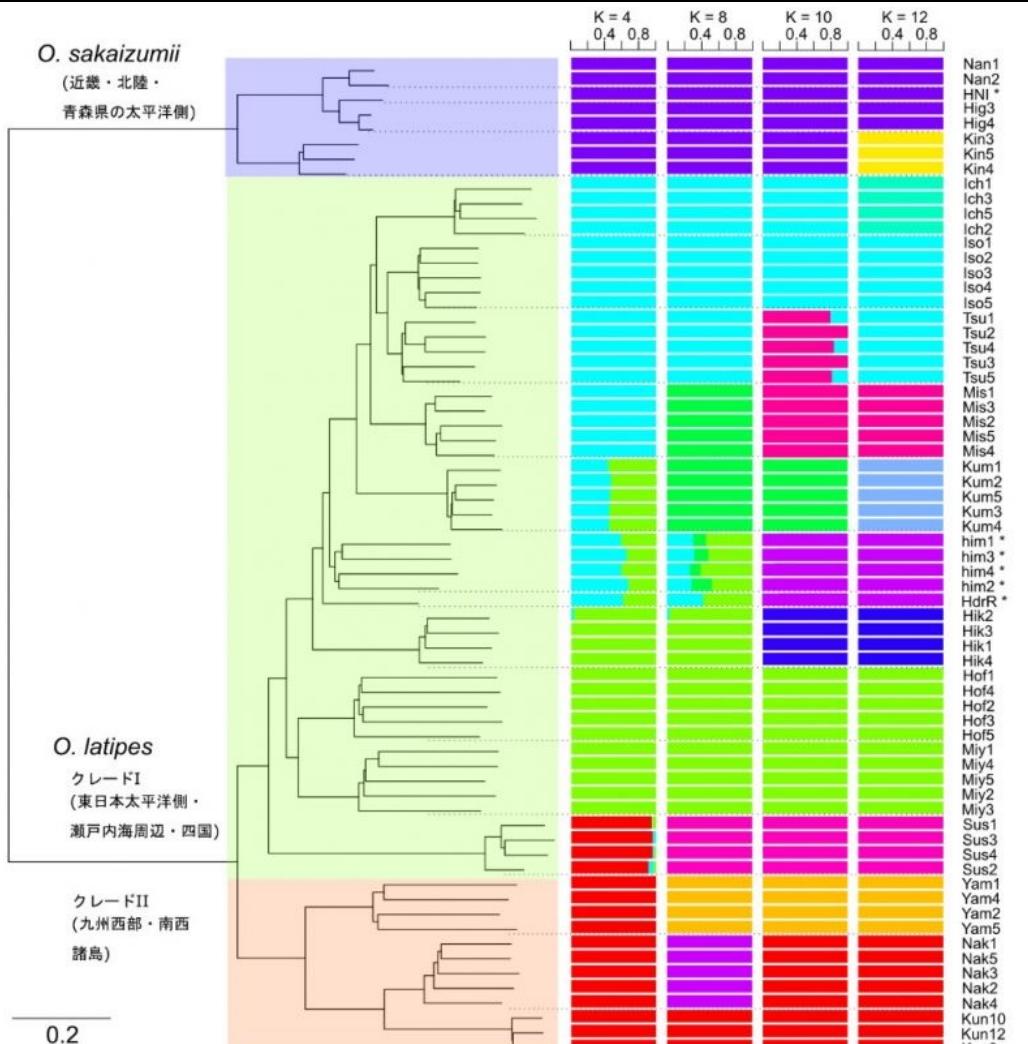
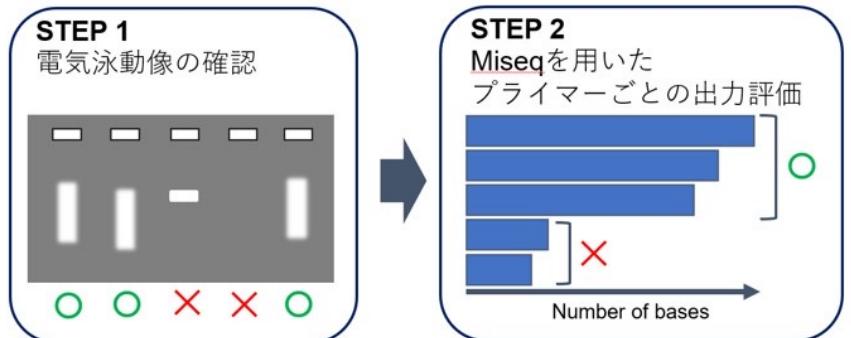
ターゲット配列の取得にランダムプライマーを用いた PCR を用いる方法。
参照ゲノム配列に依存せず、情報が得られるので簡便かつ低成本な技術として注目されている。

Developed MAAS(multiple arbitrary amplicon sequencing)

a. 計算機上のプライマー設計



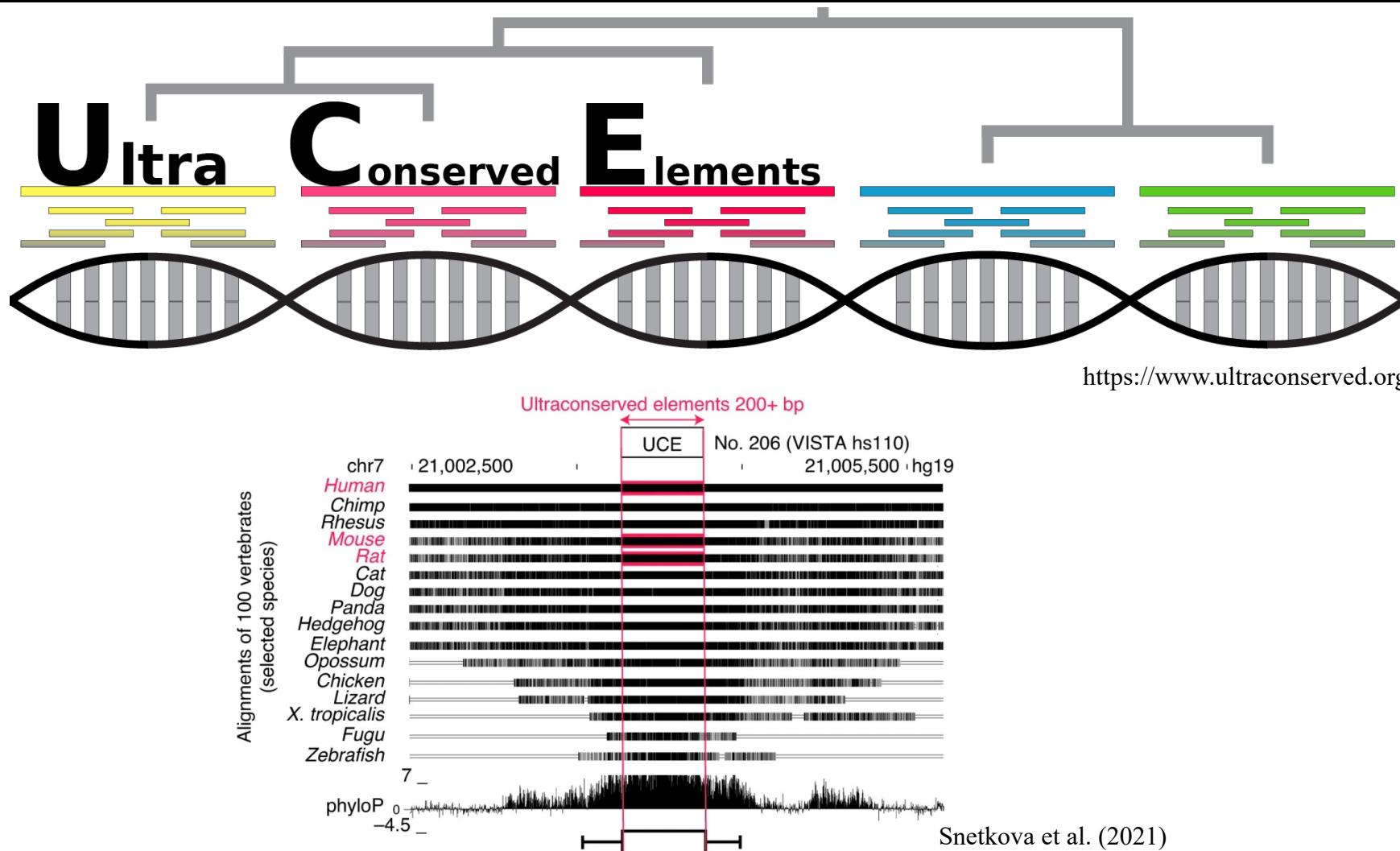
b. 分子実験による検証



Fujimoto et al. (2022)

ランダムプライマーを改善して、従来法に比べて高効率で塩基配列を決定
試したい場合は、琉球大の木村教授に相談して下さい！

UCE (ultra-conserved element) target capture



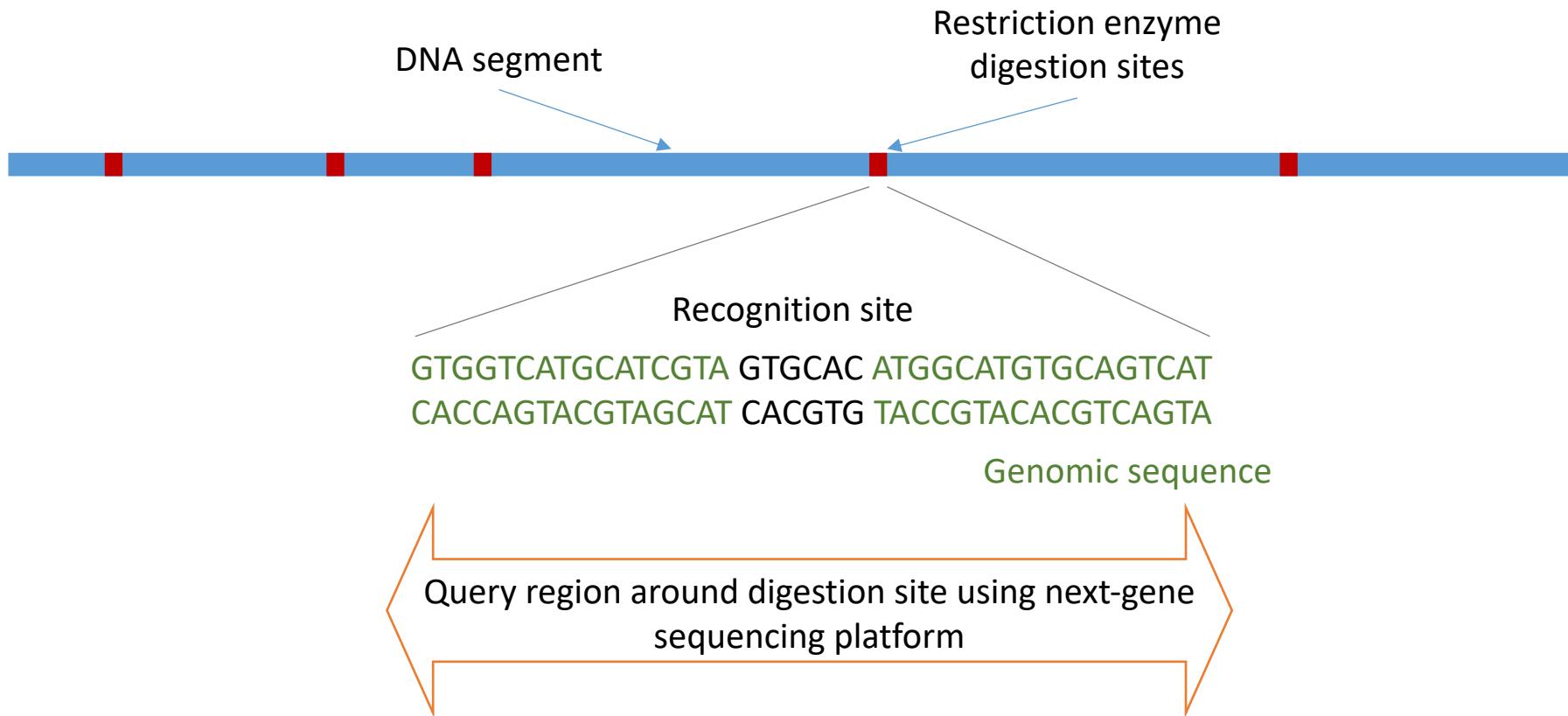
UCEは極めて高い保存性から集団ゲノミクス解析のマーカーとして活用されている。

さまざまな動物群で、UCE配列が定義されており、リファレンスゲノムがない場合でもターゲットキャプチャ法を用いることで効率よく解読することができる。

配列の保存性が高いので、集団内の遺伝的多様性よりもむしろ種間の遺伝的な違いの比較に用いられることが多い。

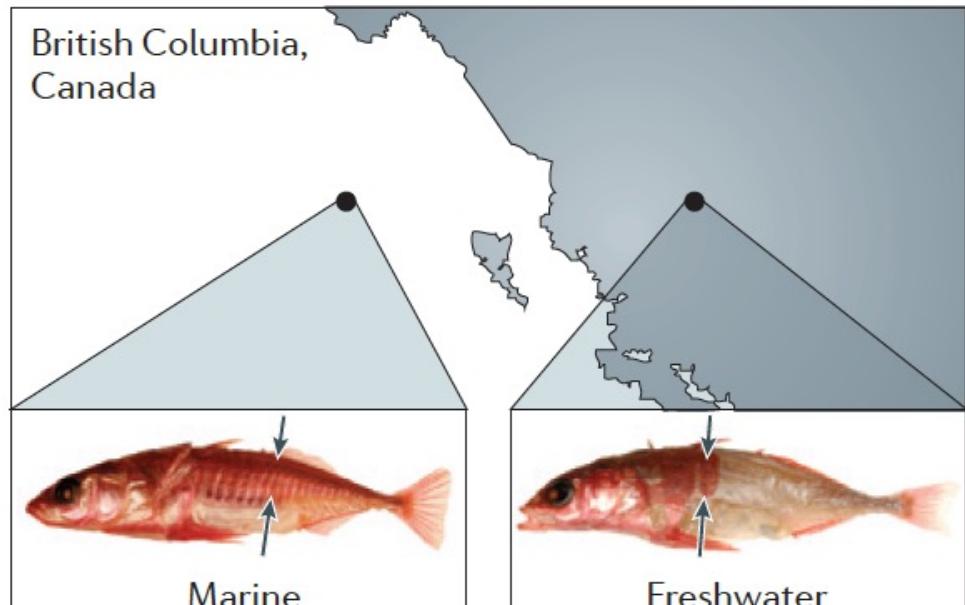
RAD-seq解析tips

RAD-seq (Restriction-site Associated DNA)

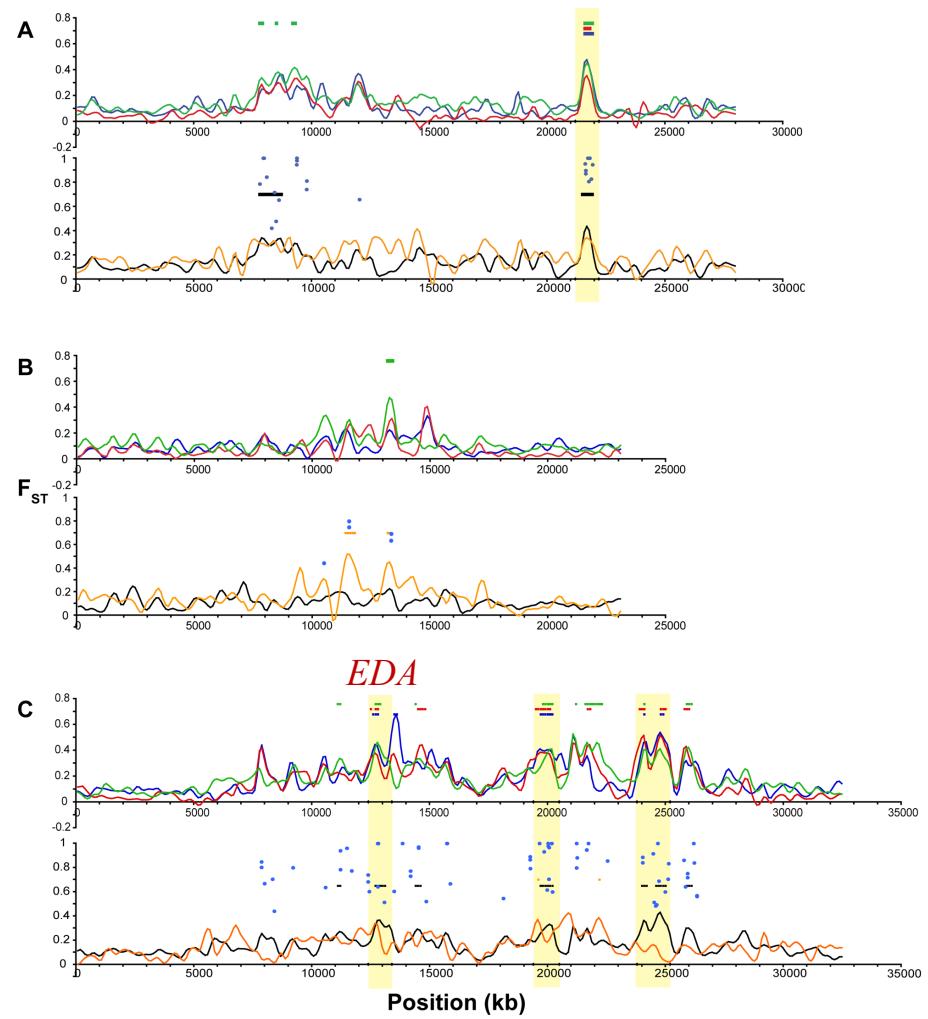


ゲノム DNA を特定の制限酵素で処理し、そのバンドパターンから遺伝的多様性を推定するAFLPを応用し、制限酵素で断片化されたゲノム DNA の部分配列をショートリードシークエンサーで解読し、そこにある SNP を検出する。
高品質のゲノムDNAが必要だが正確にSNPを検出することができる。

RAD-seqによるsticklebacks(*Gasterosteus aculeatus*)のゲノム多様性解析



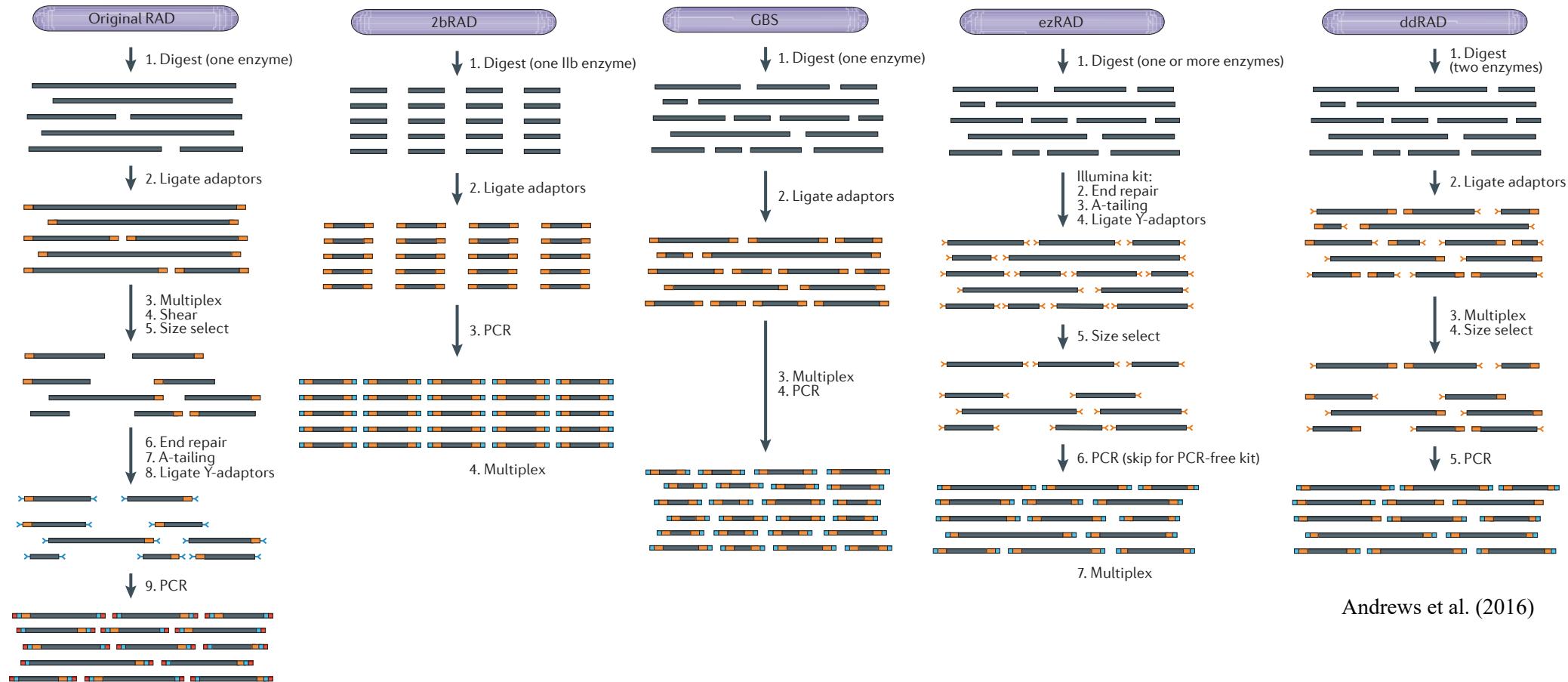
Barrett and Hoekstra (2011)



Hohenlohe et al. (2010)

RAD-seqでトゲウオの淡水型と海水型の表現型の違いの遺伝基盤の一端が明らかになった。

いろいろなRAD-seq



Andrews et al. (2016)

使用する制限酵素やサイズセレクションの方法に応じていろいろなRAD-seq解析法が提案されている。

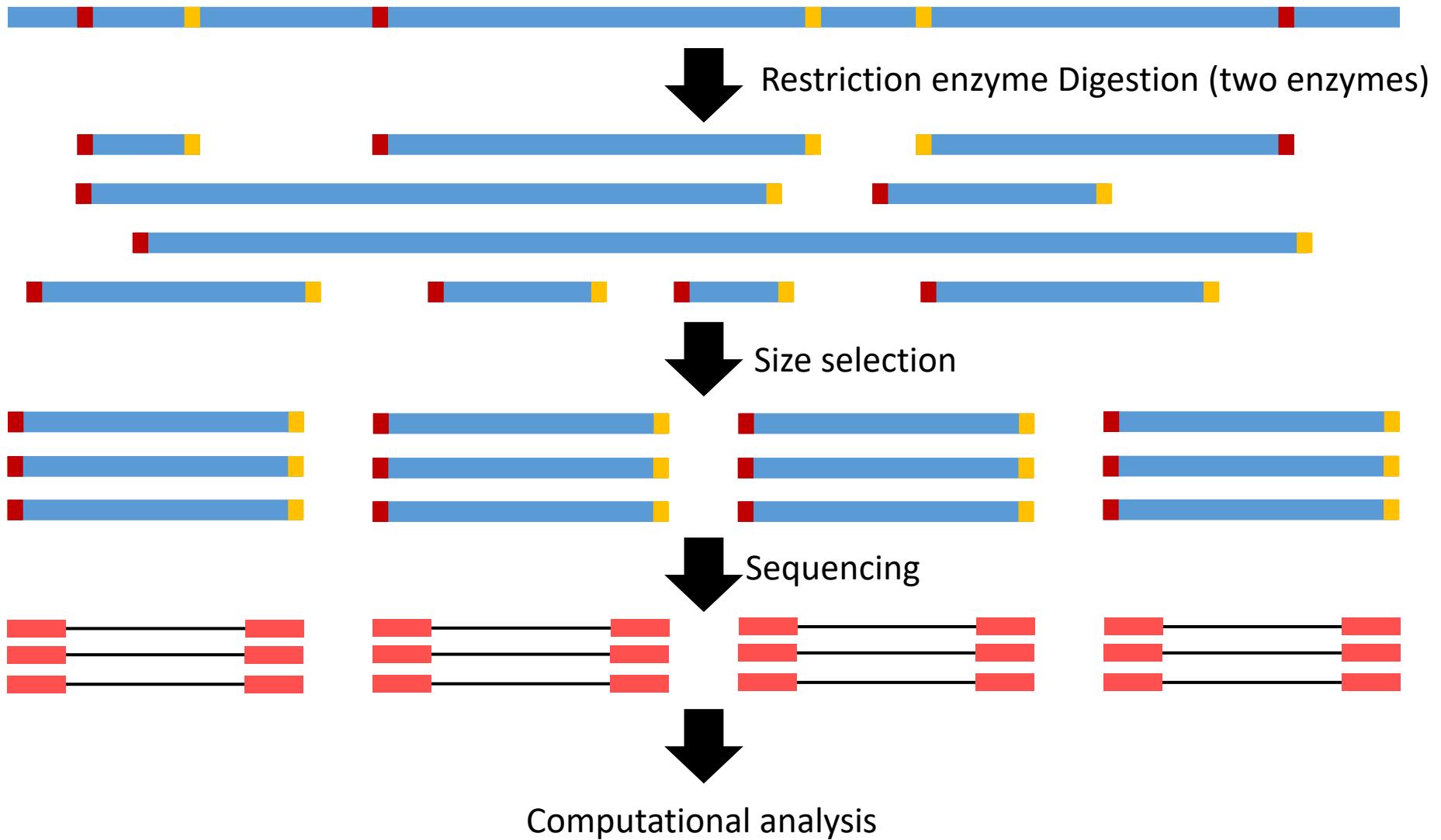
いろいろなRAD-seq

	Original RAD	2bRAD	GBS	ddRAD	ezRAD
Options for tailoring number of loci	Change restriction enzyme	Change restriction enzyme	Change restriction enzyme	Change restriction enzyme or size selection window	Change restriction enzyme or size selection window
Number of loci per 1 Mb of genome size*	30–500	50–1,000	5–40	0.3–200	10–800
Length of loci	≤1kb if building contigs; otherwise ≤300 bp [‡]	33–36 bp	<300 bp [‡]	≤300 bp [‡]	≤300 bp [‡]
Cost per barcoded or indexed sample	Low	Low	Low	Low	High
Effort per barcoded or indexed sample[§]	Medium	Low	Low	Low	High
Use of proprietary kit	No	No	No	No	Yes
Identification of PCR duplicates	With paired-end sequencing	No	With degenerate barcodes	With degenerate barcodes	No
Specialized equipment needed	Sonicator	None	None	Pippin Prep [¶]	Pippin Prep [¶]
Suitability for large or complex genomes	Good	Poor	Moderate	Good	Good
Suitability for de novo locus identification (no reference genome)*	Good	Poor	Moderate	Moderate	Moderate
Available from commercial companies	Yes	No	Yes	Yes	No

Andrews et al. (2016)

使用する制限酵素やサイズセレクションの方法に応じていろいろなRAD-seq解析法が提案されている。

ddRAD-seq (double digestion RAD)



制限酵素サイト数の見積もり方

ゲノムサイズ350Mbp, GC = 50%のとき、8bp配列を認識する制限酵素を使用すると

$$(1/4)^8 * 350,000,000 = 5,340 \text{ RAD fragments}$$

のフラグメントが得られると見積もられる。

1 siteについて x30からx60のカバレージが必要なので、x30で読むとき、NGSで必要なリード数は1個体につき、

$$10,680 * 30 = 320,440 \text{ (0.32 M read)}$$

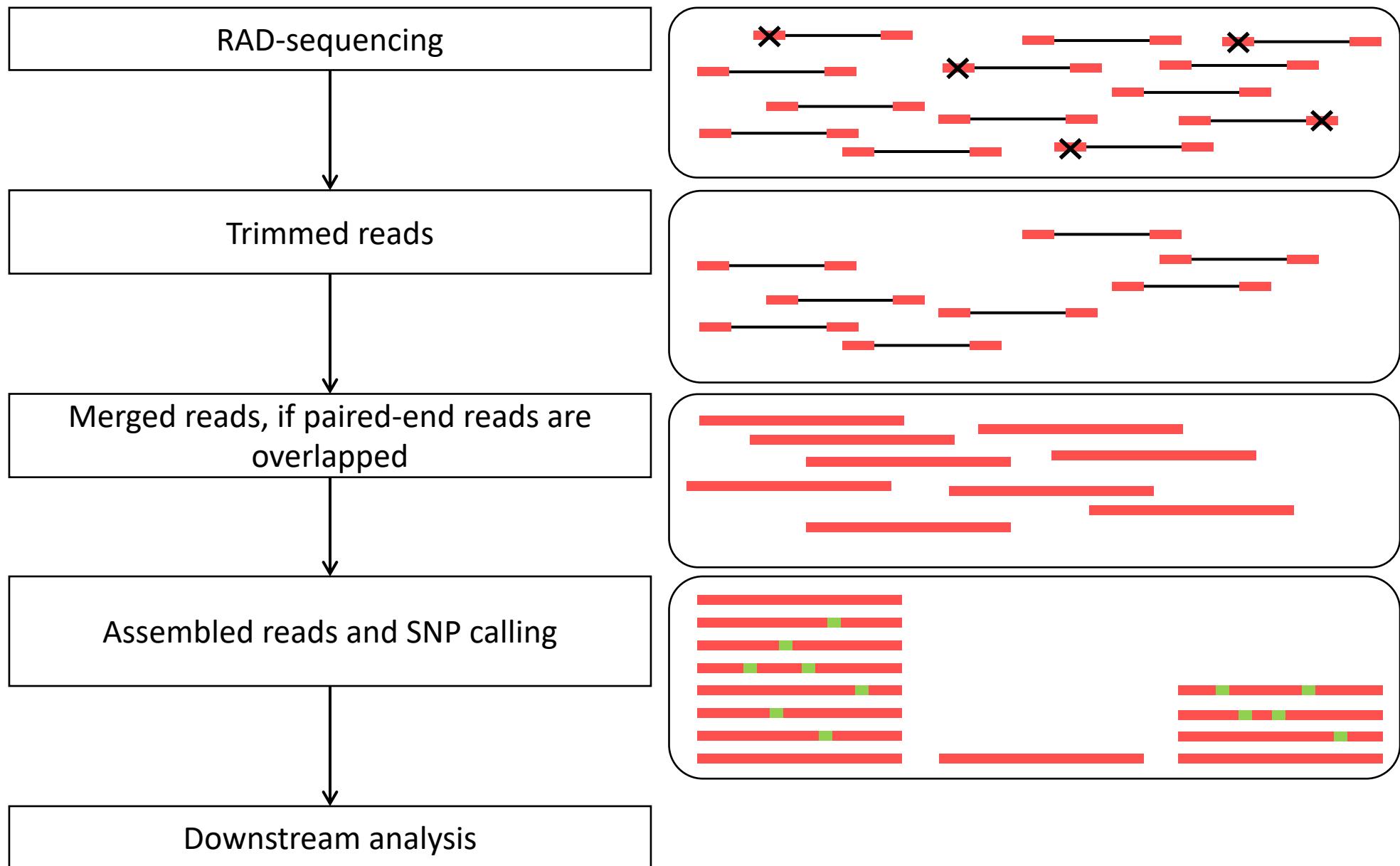
となる。

ちなみにゲノムサイズ16Gbp, GC = 50%のだと、15 M read必要

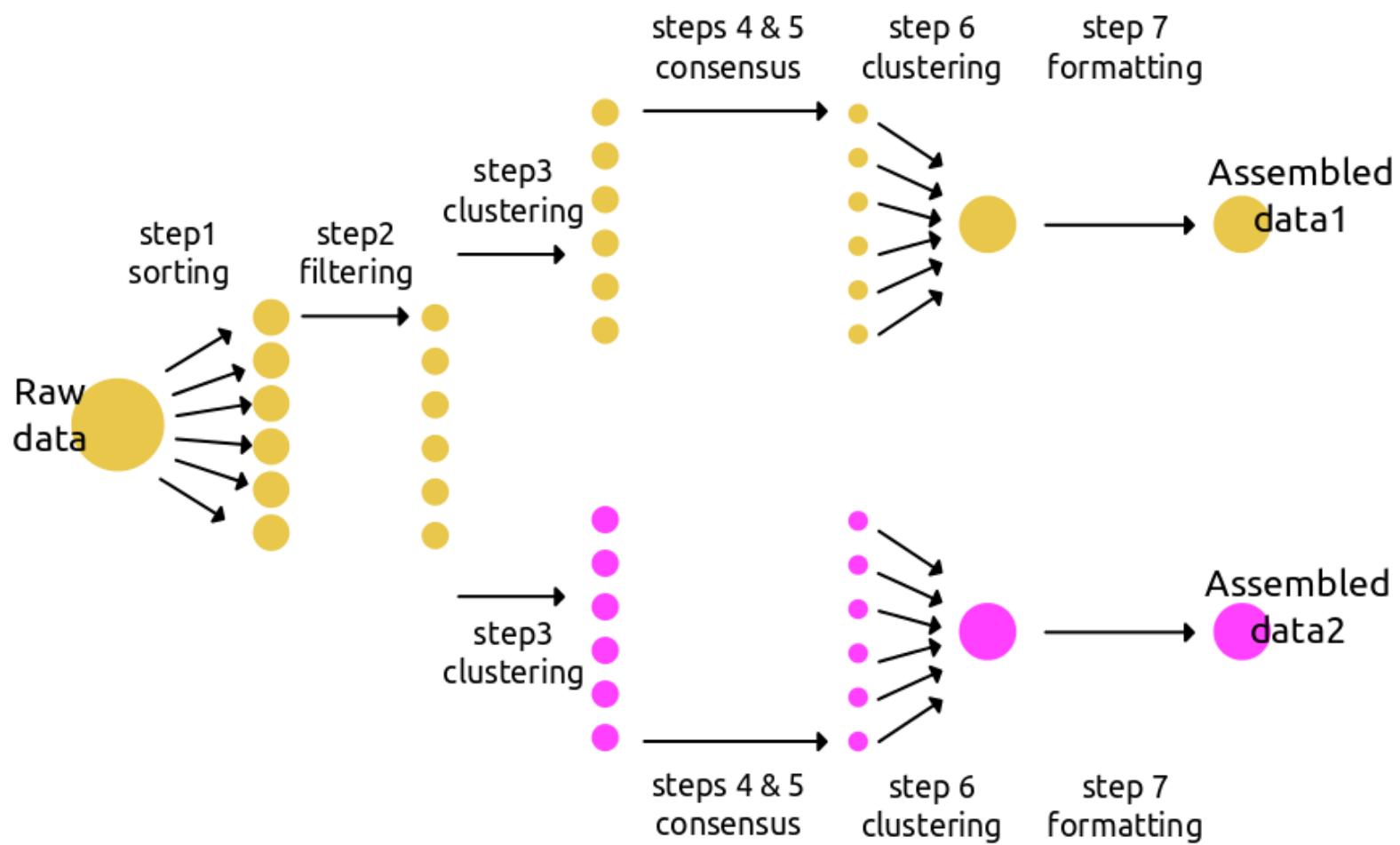
ゲノムが解読されていれば、このような計算は必要なくゲノムから調べることができる。

=> 経験的には、計算通りにいかないことが多い。

Pipeline of RAD-seq analysis (de novo)



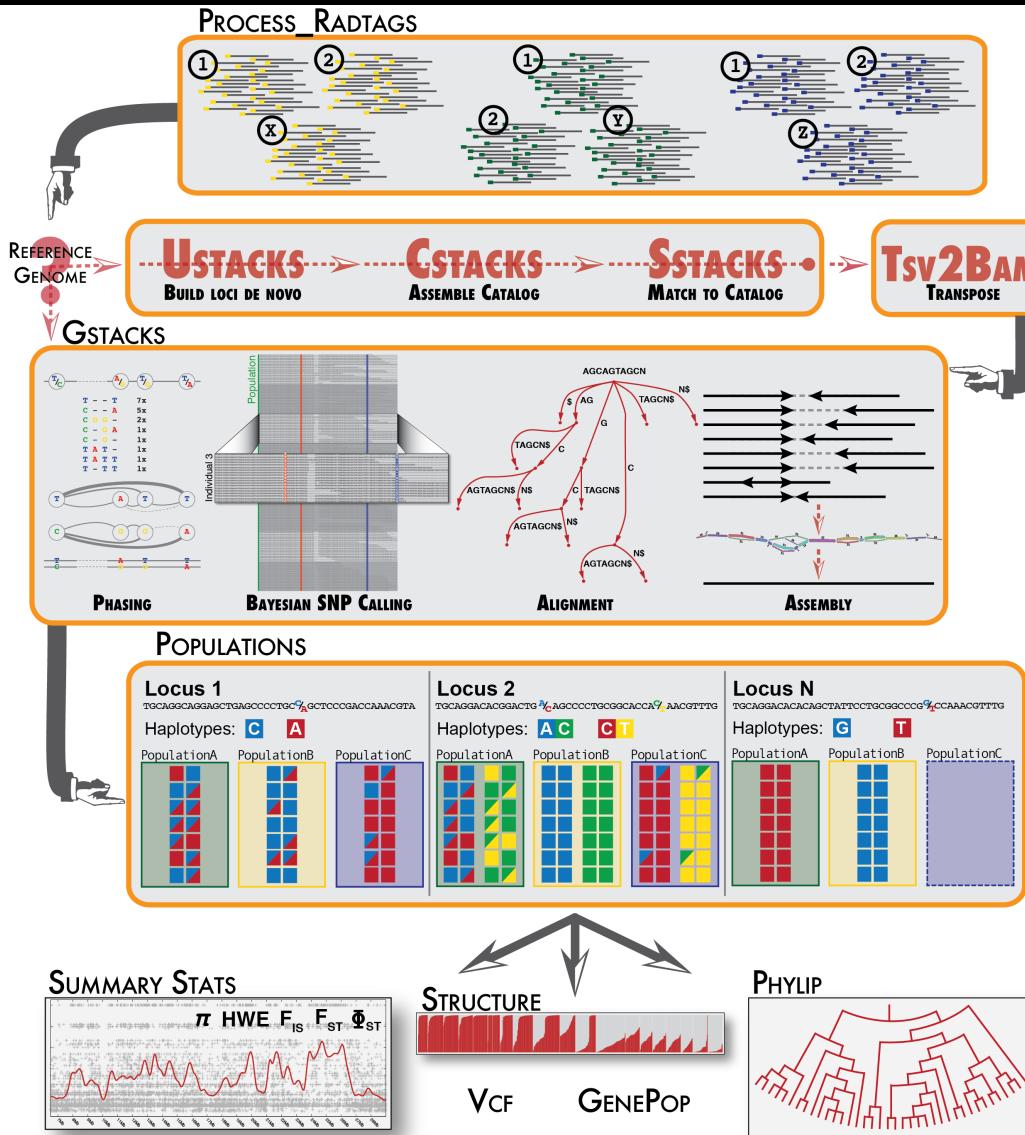
RAD-seq解析でよく使われるprogram: ipyrad



<https://ipyrad.readthedocs.io/en/latest/>

植物ゲノム研究者が開発。Merged readにも対応。

RAD-seq解析でよく使われるprogram: Stacks



<http://catchenlab.life.illinois.edu/stacks/manual/>

Readの長さが統一されていないと解析できない。

RAD-seq解析の注意点①:制限酵素サイトにあるSNP

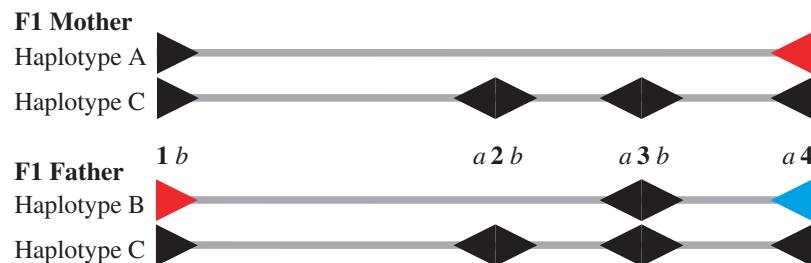


Fig. 5 Examples of complex RAD loci. Four consecutive restriction sites are shown (1–4) in each of four haplotypes for two F1 parents, with RAD loci on either side of each site shown as arrowheads and labelled a and b. Haplotype C is shared between the parents. Variant alleles at RAD loci are differentiated by colour (black, red and blue). For example, two variant alleles are found at RAD locus 1b, with the black allele shared between haplotypes A and C and the red allele present in haplotype B only; the black allele has two copies in the mother and one copy in the father. Owing to the heterozygous restriction sites 2 and 3, alleles at RAD loci 1b and 4a are derived from restriction fragments of varying lengths, potentially skewing read depths and resulting in incorrect genotypes. For example, at locus 1b, the father's red allele at Haplotype B will be derived from a long fragment ending at RAD locus 3a, whereas the black allele at Haplotype C will be derived from a shorter fragment ending at RAD locus 2a.

Davey et al. (2013)

制限酵素サイトにSNPがあると、サンプルごとに
読まれるゲノム領域が変わり、解析結果に影響
する。

RAD-seq解析の注意点②:エラー

Source	Reason	Reference	Source	Reason	Reference
Technical and human error			Paralogs and repetitive regions	Paralogous and repetitive regions with similar sequences can be erroneously merged together as a single locus	Hohenlohe <i>et al.</i> (2012); Dou <i>et al.</i> (2012)
Technical	Errors related inversely to the quality of reagents and equipment, and to the organization of the laboratory in different rooms to avoid contamination	Bonin <i>et al.</i> (2004)	Presence of indels	Stacks and RADtools are unable to handle indels; therefore, indel-containing loci are not clustered together, while they can be recovered by RaPID and pyRAD	Peterson <i>et al.</i> (2012); Davey <i>et al.</i> (2013)
Human	Sample mislabelling, sample contamination, pipetting error and error during DNA concentration measurements	Bonin <i>et al.</i> (2004)	Mapping using a reference genome	Mapping of alleles that are different from the reference genome is less probable than for a reference-matching allele, causing a bias in allele frequency towards the allele found in the reference sequence. It may additionally reduce the number of SNPs discovered and bias estimates of nucleotide diversity towards smaller values	Pool <i>et al.</i> (2010)
Wet laboratory					Mastretta-Yanes <i>et al.</i> (2014)
Enzyme sensitivity to DNA quality and quantity	Digestion and PCR efficiency may be uneven among samples, which can result in the underrepresentation of some restriction fragments	Bonin <i>et al.</i> (2004)			
Pooling concentration	Samples with higher concentration can be overrepresented in the sequencing output if they are not pooled in equimolar amounts	Baird <i>et al.</i> (2008); Peterson <i>et al.</i> (2012)			
PCR error	PCR error may get further amplified and can appear in multiple reads resembling an alternative allele at a locus. PCR error may differ among samples depending on reaction conditions and experimental design	Hohenlohe <i>et al.</i> (2012)			
PCR bias	PCR amplification success may be variable across different alleles or barcodes, biasing their representation. Differences in amplification success lead to variation of coverage among loci and individuals, potentially resulting in allelic dropout, nonrepresentation of some loci or PCR duplicates	Bonin <i>et al.</i> (2004); Peterson <i>et al.</i> (2012); Hohenlohe <i>et al.</i> (2012)			
Size selection (double digest)	Different fragments may be selected if more than one excision is performed. Imprecise size selection can include fragments of lengths relatively distant from the size-selection target mean	Peterson <i>et al.</i> (2012)			
Exposure to UV light	Can produce fragmentation (that could lead to locus/allele dropouts) and mutation of DNA strands (that introduces nonbiological variation)	Grundemann & Schomig (1996)			
Next-generation sequencing (NGS)					
Sequencing error	NGS introduces sequencing error (0.1–1.0% per nucleotide) that can vary across samples, restriction site-associated DNA (RAD) sites and positions in the reads for each site	Hohenlohe <i>et al.</i> (2012), Meacham <i>et al.</i> (2011); Nielsen <i>et al.</i> (2011); Loman <i>et al.</i> (2012)			
Sequencing sampling	The sampling process of a heterogeneous library inherent in NGS introduces sampling variation in the number of reads observed across RAD sites as well as between alleles at a single site	Hohenlohe <i>et al.</i> (2012)			
Barcode error	PCR or sequencing errors at the DNA-tag of a fragment can reduce the number of reads obtained for it	Hohenlohe <i>et al.</i> (2012)			
Genome intrinsic GC content	At large numbers of PCR cycles, RAD loci with high GC content are sequenced at higher depths compared to RAD loci with low GC content. But at the same time, high GC content loci could be undersequenced if too few PCR cycles are performed. GC bias contributes to PCR duplicates	Davey <i>et al.</i> (2013)			
Restriction site variation	Variation in the restriction site within a locus will result in allelic dropout	Davey <i>et al.</i> (2013); Gautier <i>et al.</i> (2013b)			
DNA methylation	For some restriction enzymes, digestion is impaired or blocked by methylated DNA. The same gene may or may not be methylated in different individuals or tissues	Roberts <i>et al.</i> (2010)			
Bioinformatic					
Variation in coverage	Coverage is an important filter to distinguish real variation from sequencing errors, repetitive regions and duplicates. But if there is coverage heterogeneity among samples and alleles, or if the general coverage is low, setting the filters with minimal coverage values too high can lead to allelic dropout. Setting it too low, however, can lead to incorrect single-nucleotide polymorphism (SNP) calls	Hohenlohe <i>et al.</i> (2012); Davey <i>et al.</i> (2013); Catchen <i>et al.</i> (2013)			
PCR duplicates	PCR duplicates occur when more than one copy of the same original DNA molecule attaches to different beads/cells during sequencing. This can result in high coverage of PCR error variation, or it can produce heterogeneous coverage distribution due to GC and PCR bias	Davey <i>et al.</i> (2013)			
Fragment length	Alleles will drop out as restriction fragment length decreases because RAD loci from short restriction fragments have low read depths. The efficacy of different bioinformatics tools at dealing with this varies	Davey <i>et al.</i> (2013)			

いろいろとシークエンスエラーが生じる可能性がある。

RAD-seq解析の注意点③: 最適化が必要なパラメーター

どれだけのカバレージで読んでいるリードを用いるかを指定するパラメータ m

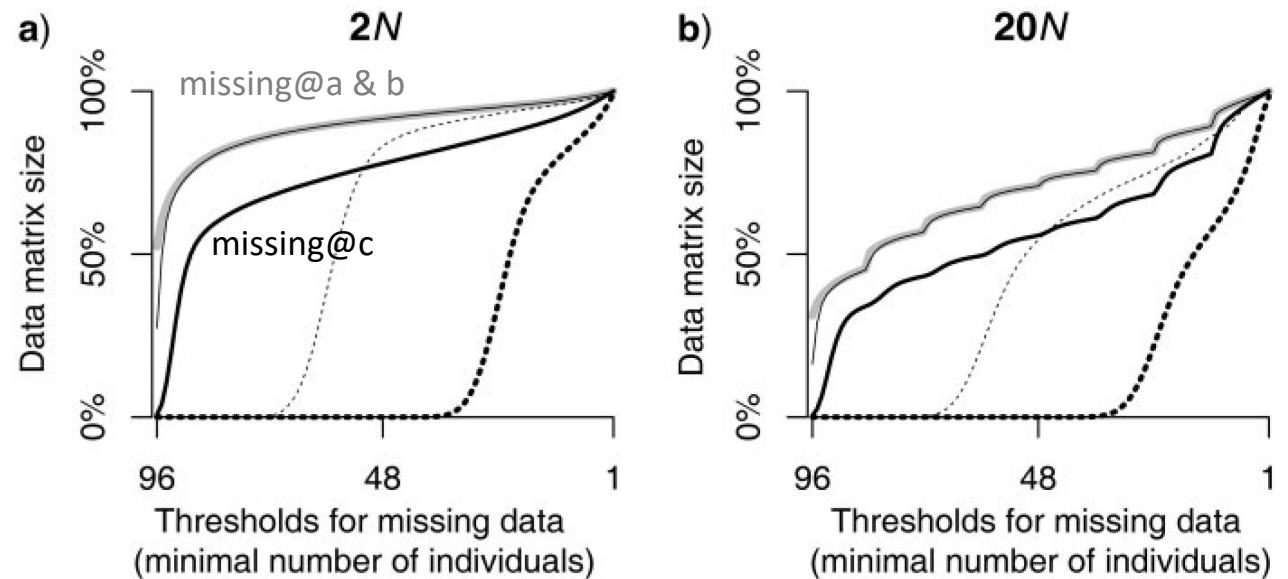
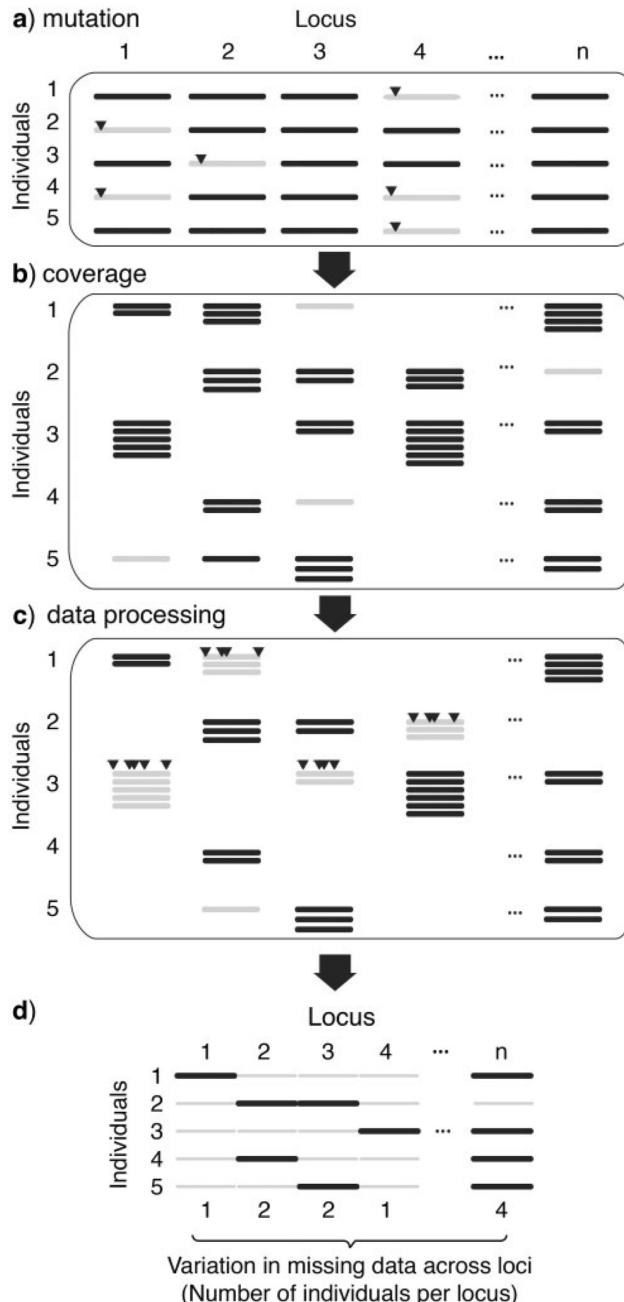
クラスタリングの際にリード間でどれだけのミスマッチを許すかを指定するパラメータ M

カタログを作る際にミスマッチをいくつ許すかを指定するパラメータ n

Parameter	How it affects assembly and genotyping error*
Minimum number of identical, raw reads required to create a stack (- m) default 3	Reads with convergent sequencing errors are probably to be erroneously labelled as stacks if $-m$ is too low. True alleles will not be recorded and will drop out if $-m$ is too high. $-m$ can decrease genotyping error by distinguishing real loci from PCR and sequencing error, but it can increase error by calling a heterozygous locus as homozygous when minimum coverage is set too high and one of the alleles is therefore excluded
Number of mismatches allowed between loci when processing a single individual (- M) default 2	If $-M$ is too low, some real loci will not be formed, and their alleles will be treated as different loci (undermerging). If $-M$ is too large, repetitive sequences and paralogs will form large nonsensical loci (overmerging)
Number of mismatches allowed between loci when building the catalogue (- n) default 0	For $n = 0$, there would be loci represented independently across individuals that are actually alleles of the same locus. If $n > 0$, the consensus sequence from each locus is used to attempt to merge loci. This is important for population studies where monomorphic or fixed loci may exist in different individuals. Merging fixed alleles as a single locus can increase the probability of assembling real loci and therefore decrease the allele error rate. However, erroneous loci will be created if $-n$ is too high
Maximum number of stacks at a single <i>de novo</i> locus (--max_locus_stacks) default 3	The expectation for nonrepetitive genomic regions is that a monomorphic locus will produce a single stack because the two sequences on the two homologous chromosomes are identical and thus indistinguishable. In contrast, a polymorphic locus will produce two stacks representing alternative alleles. Confounding cases that may arise from short, sequencing error-based stacks or from repetitive sequences, where hundreds of loci in the genome may collapse to a single putative locus. --max_locus_stacks allows for the identification and blacklisting of confounding cases
Single-nucleotide polymorphism (SNP) calling model	In the default SNP calling model, the error parameter is allowed to vary freely, whilst in a bounded-error model, the boundary value is substituted if the maximum-likelihood value of ϵ exceeds a lower or upper bound. One consequence is that reducing the upper bound increases the chance a homozygous loci being called heterozygous. The SNP calling model allows the tolerance for false positive vs. false negative rates in calling genotypes to be tuned, which in turn influences the genotyping error

*Parameters explanation as in Catchen *et al.* (2013) and *Stacks* documentation, effect on genotyping error as discussed here.

RAD-seq解析の注意点④: missing data



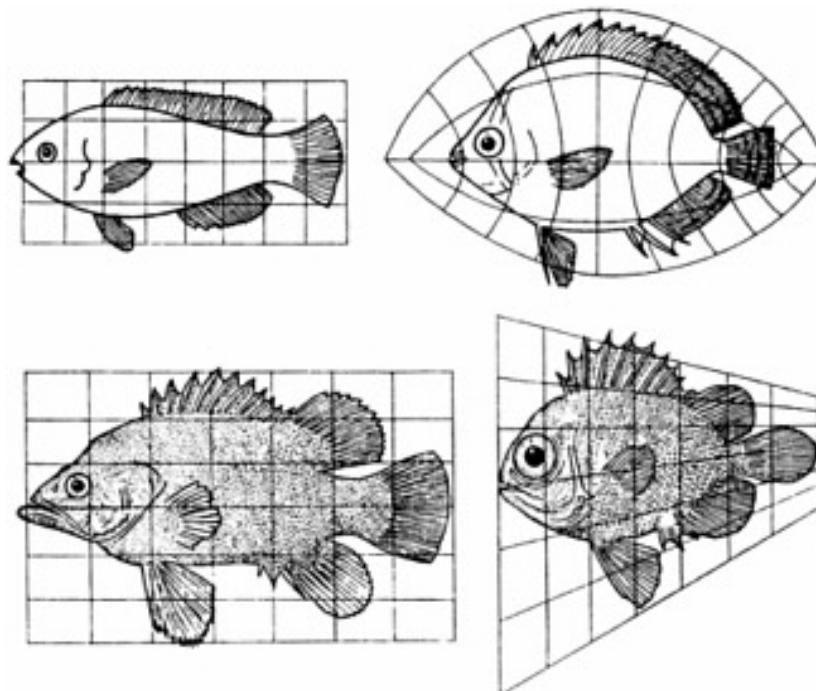
Huang et al. (2014)

解析のどの過程でmissing dataが生じるかによって結果に影響する。

実際に解析する際は、時間が許すようなら個々のパラメータの値を検討して、安定した結果が得られる最適な設定を探求すべき。

実例紹介

幾何学的形態測定法 (Geometric morphometrics)



Thompson (1917)

Geometric morphometrics is the quantitative analysis of form, a concept that encompasses size and form.

分子生態学の分野では、遺伝情報と形質情報を統合して解析するのが一般的です。形態の表現型解析には幾何学的形態測定法を用いることが多いです。

MedakaのddRAD解析 (reference genomeあり)

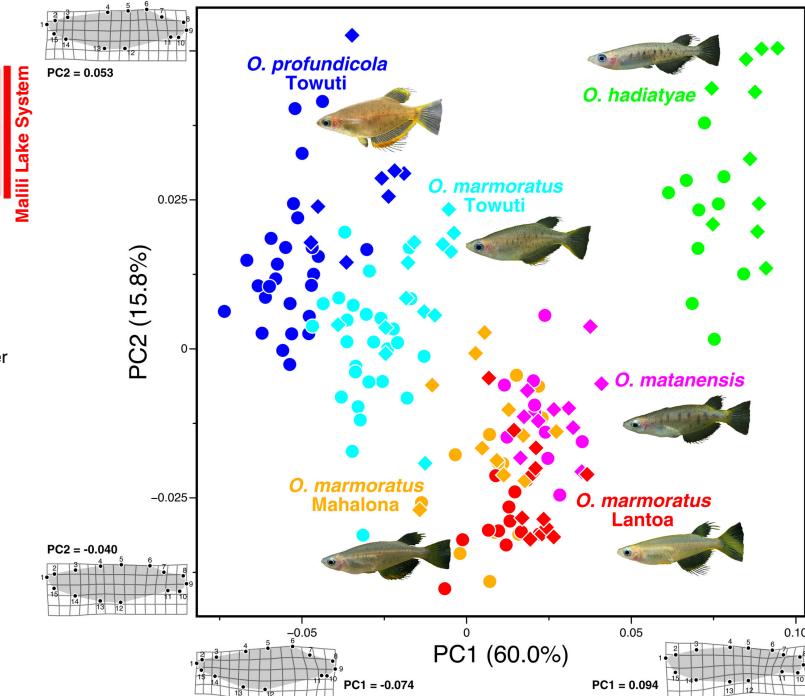
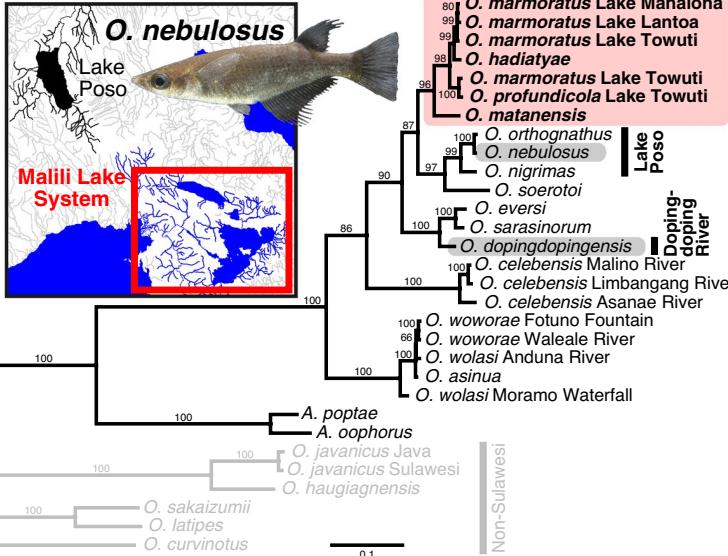
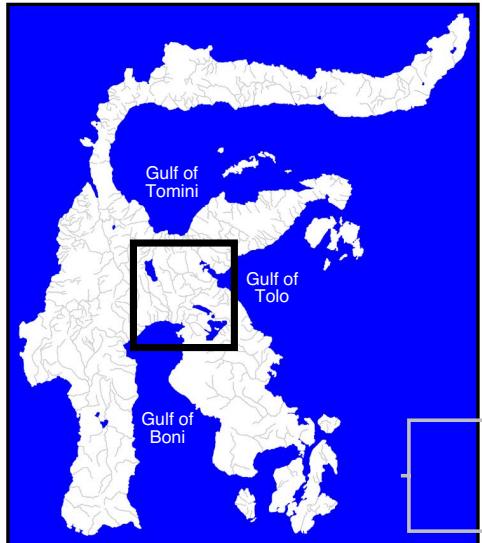
Received: 17 September 2020 | Accepted: 9 September 2021
DOI: 10.1111/jeb.13932

RESEARCH ARTICLE

JOURNAL OF EVOLUTIONARY BIOLOGY WILEY

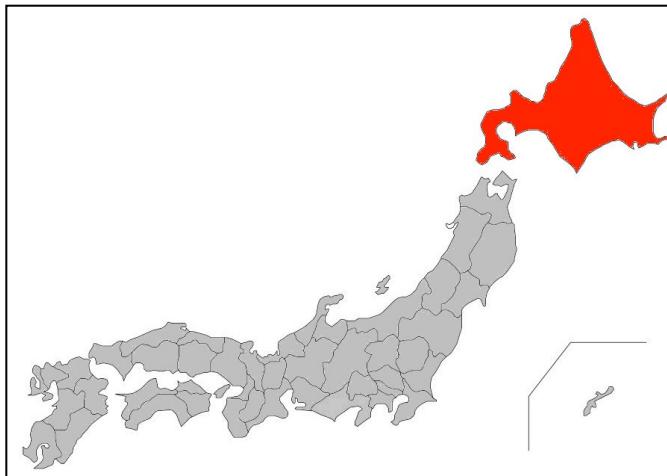
Species divergence and repeated ancient hybridization in a Sulawesian lake system

Ixchel F. Mandagi^{1,2} | Ryo Kakioka¹ | Javier Montenegro¹ | Hirozumi Kobayashi¹ | Kawilarang W. A. Masengi² | Nobuyuki Inomata³ | Atsushi J. Nagano^{4,5} | Atsushi Toyoda⁶ | Satoshi Ansai⁷ | Masatoshi Matsunami⁸ | Ryosuke Kimura⁸ | Jun Kitano⁹ | Junko Kusumi¹⁰ | Kazunori Yamahira¹



This resulted in 2,252 RAD loci, which included 2,008 polymorphic loci.

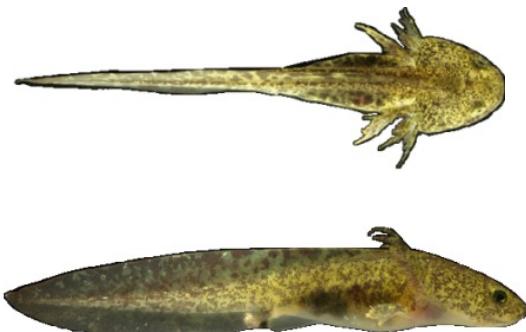
エゾサンショウウオ (*Hynobius retardatus*)



- 生息地: 北海道各地
- 体長: 11-19 cm (成体), 2-3cm (幼生)
- 生活史: 4-6月頃雪解け水でできた水たまりで孵化し、約60日で成体に変態する。
- 幼生は生息環境に応じた表現型可塑性を示す。

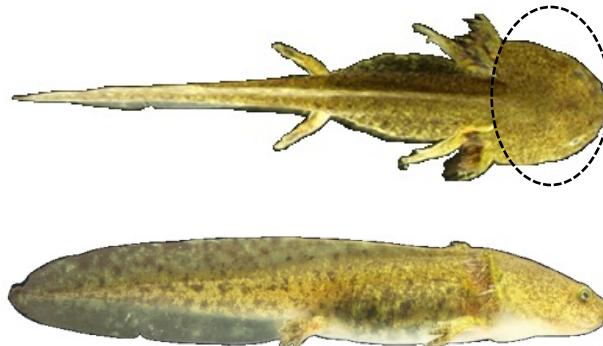
エゾアカガエル幼生(オタマジャクシ)によって 誘導される表現型可塑性

基本型



攻撃型

頭部が発達

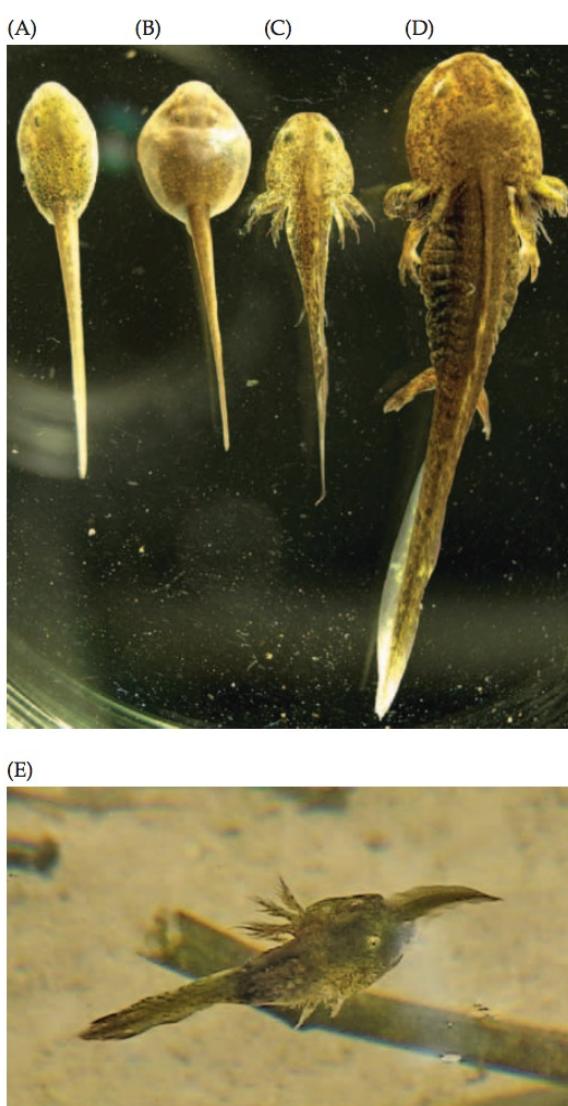


エゾアカガエル幼生と共に存させてサンショウウオを飼育すると、オタマジャクシの捕食に有利な頭でっかちな攻撃型の表現型が誘導される。

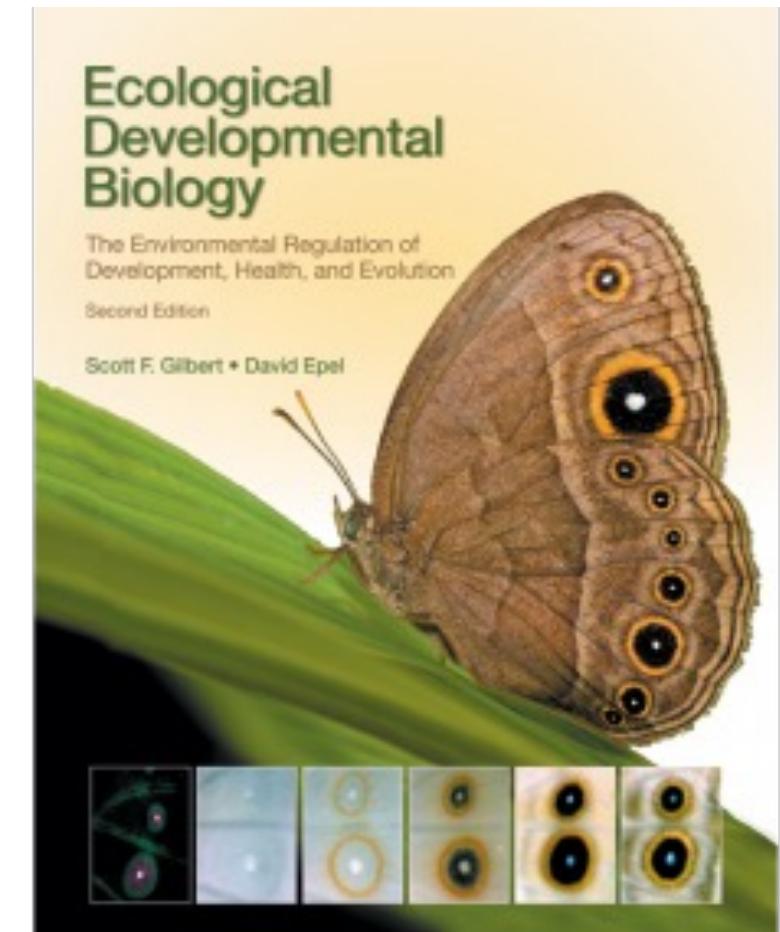
Creating Each Other's Environment: Reciprocal Developmental Plasticity

Niche construction results when the traits of an organism influence the habitat that it experiences. Sometimes this can result in strange consequences. In Chapter 1, we discussed predator-induced polyphenism, where an organism's plasticity enables it to respond to a predator (developmentally, physiologically, or behaviorally) by becoming less readily eaten. However, if this strategy is successful, one would expect that the predator should adapt to it. If the predator can't evolve a strategy to circumvent this new adaptation in the prey, it must find another food source or it will die. This is sometimes called "the evolutionary arms race." Here, predator-induced plasticity can be coupled with prey-induced plasticity in the predator.

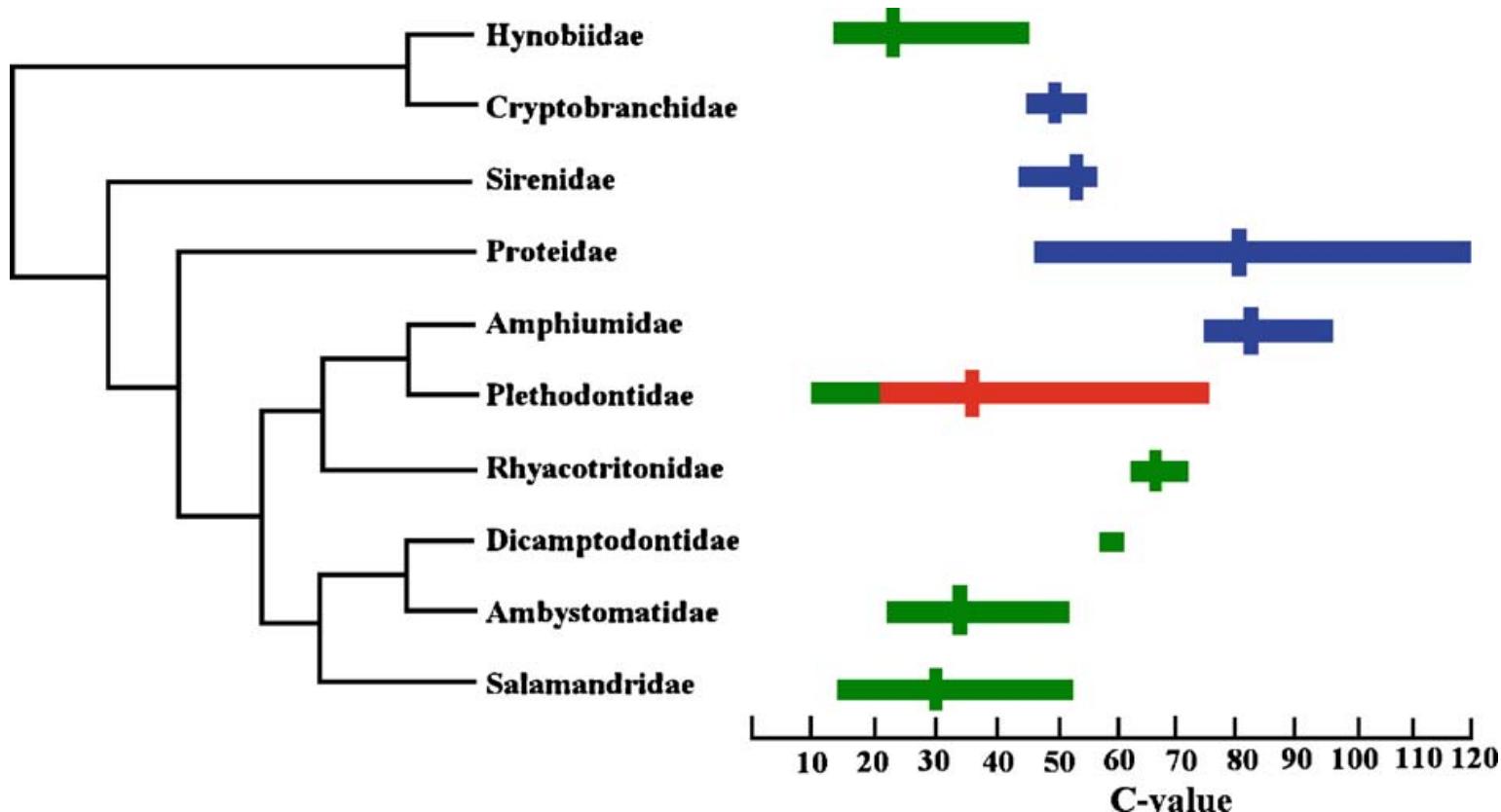
Defensive structures are often difficult to make, which is why their induction only in the continued presence of predators is such a good strategy. In northern Japan, the melting snow forms ephemeral ponds in which both salamanders (*Hynobius retardatus*) and frogs (*Rana pirica*) lay their eggs. Thus, the salamander larvae and the tadpoles (frog larvae) inhabit the ponds in early spring (Figure A–D). The salamander larvae are carnivores and will consume the tadpoles. When such salamander predators are present, they unknowingly signal the tadpoles to develop a "bulgy" phenotype with enlarged bodies and tails. This prevents their immediate ingestion by the salamanders. Conversely, when the large tadpoles are abundant, they inadvertently signal the salamander larvae to develop a broad, gaping mouth, which allows them to swallow tadpoles more easily (Figure E; Michimnae and Wakahara 2002; Kishida et al. 2006). Thus, both predator and prey use developmental plasticity to increase their fitness. Moreover, the timing of salamander larva metamorphosis is also plastic and is influenced by the number and morphology of the prey. This causes strange population fluctuations that depend on how rapid the speed of adaptation is between predator and prey (Mougi 2012; Kishida et al. 2013). In this way, there is a mutual inductive response, wherein each organism changes the environment of the other.



Reciprocal plasticity. (A–D) Offensive and defensive morphs of the salamander larvae (*Hynobius retardatus*) and frog larvae (*Rana pirica*) that inhabit the same ephemeral ponds. (A) Nondefensive frog tadpole. (B) Defensive "bulgy" frog tadpole. (C) Nonpredatory salamander larva. (D) Predaceous salamander larva. (E) Salamander larva consuming frog tadpole in natural pond. (From Dr. Kishida Osamu, forestcsv.ees.hokudai.ac.jp/en/research.)



Problem: 有尾両生類のゲノムは巨大

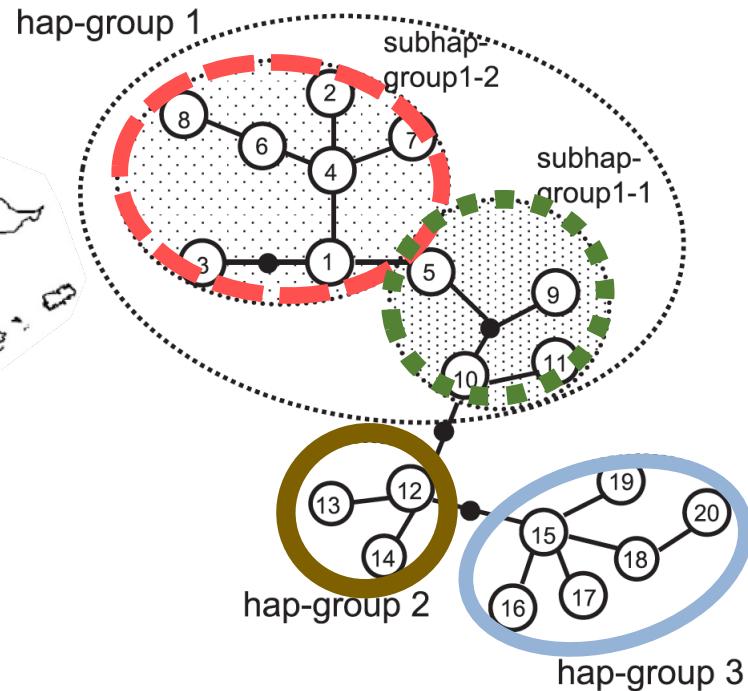
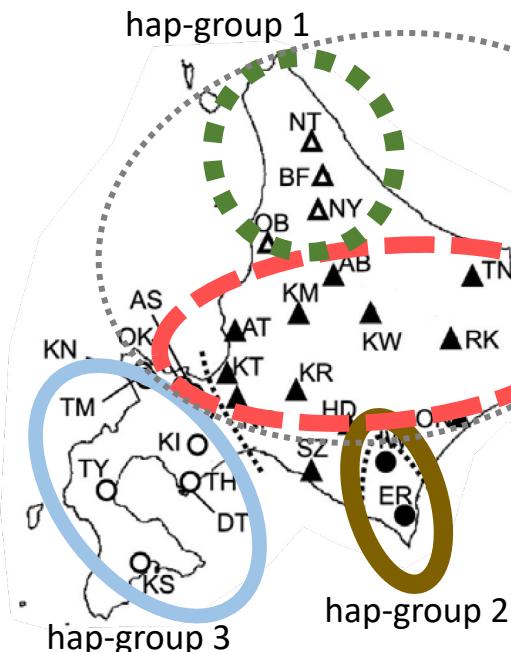


Sessions et al. (2008)

ヒトのC-valueは、約3.00なので、その7倍以上！
しかし、とりあえずddRAD-seqしてみた。

mtDNAによる系統関係

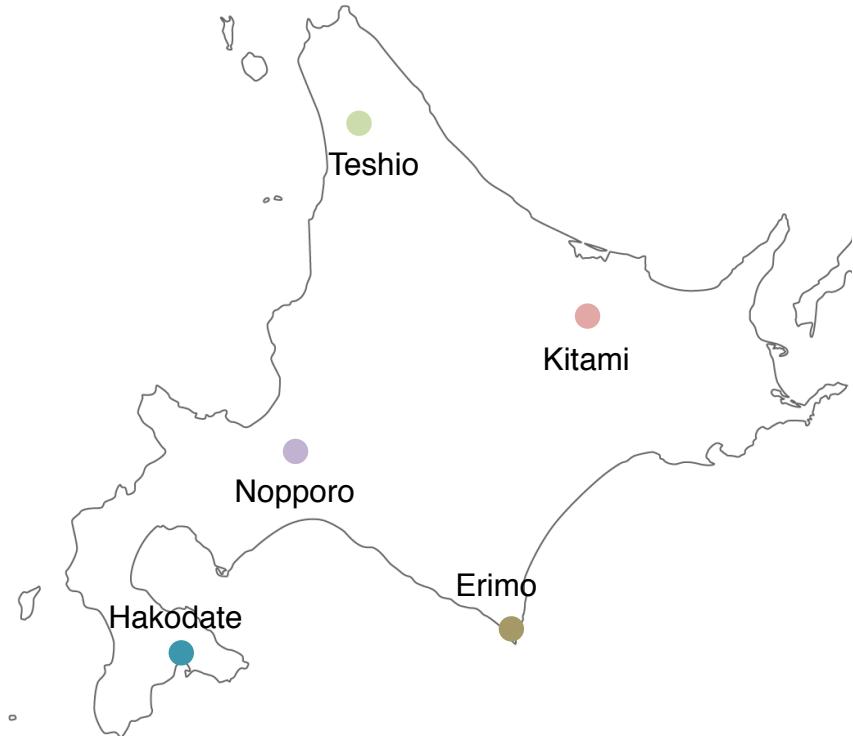
mtDNA haplotype network



Azuma et al. (2013)

mtDNAのControl領域 (490 bp)による解析からエゾサンショウウオには3つのハプロタイプがあることが明らかになっている。

道内5箇所からのサンプリングと可塑性誘導実験



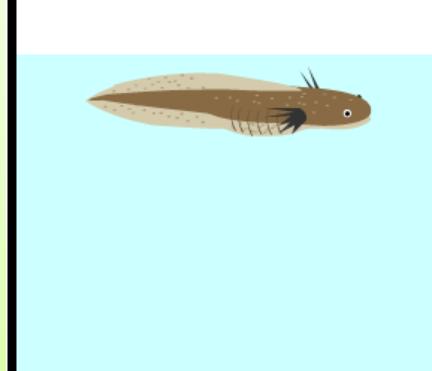
Sampling points

孵化(Day-0)

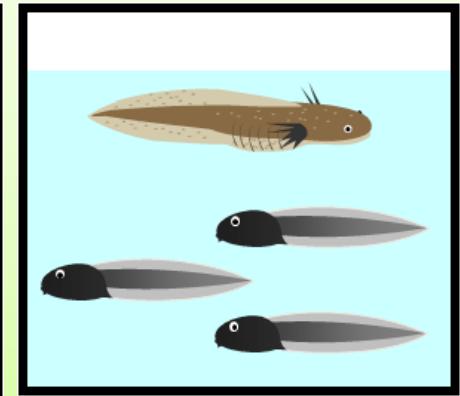
↓
実験開始

(Day-5: measured)

Tadpole control



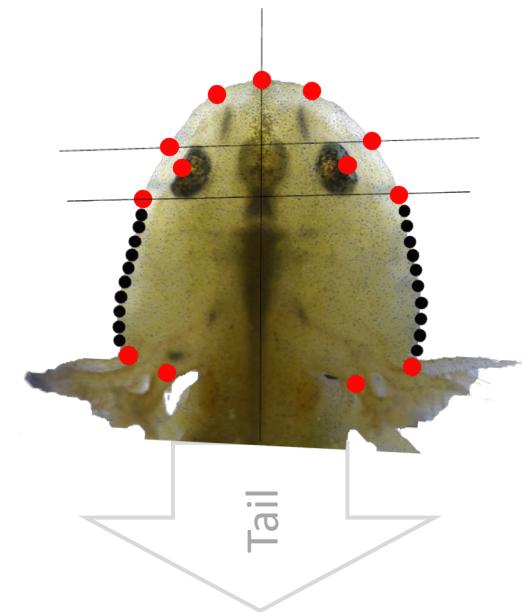
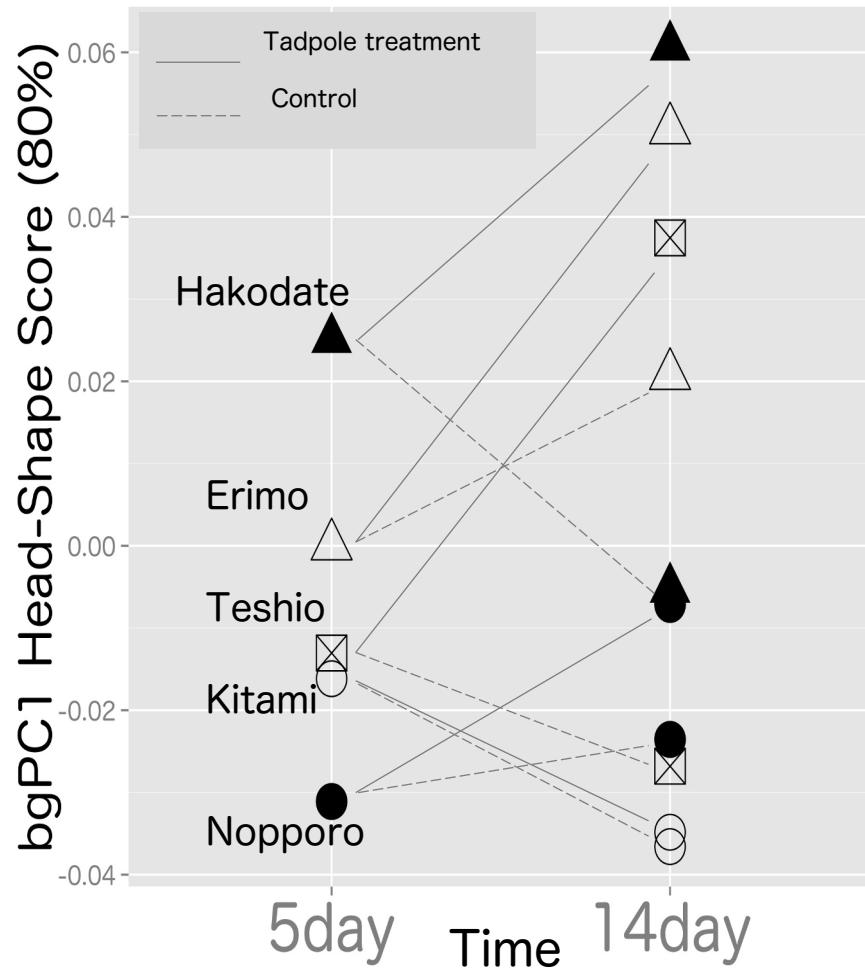
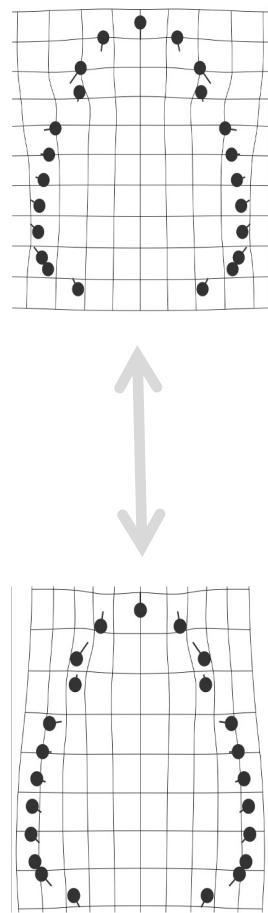
Tadpole treatment



↓
実験完了

(Day-14: measured)

攻撃型の可塑性は地域集団によって異なる



Matsunami et al. *under review*

遺伝的な違いか？環境による違いか？

Result: シーケンシング

- 100個体(各集団20個体)から、計約18G reads (150 bp paired-end reads)を取得した。
- 特に1 M reads 以下のサンプル(函館1個体, 北見2個体, 野幌1個体)は、データが少ないため解析から除いた。

	No. of sample	Average (reads/ sample)	Average after trimming (reads/ sample)	Maximum (reads)	Minimum (reads)
Erimo	20	3,800,488	3,076,824	9,764,697	1,071,031
Hakodate	20	3,150,393	2,425,669	10,503,169	593,809
Kitami	20	3,080,889	2,530,569	4,483,878	654,977
Nopporo	20	5,153,448	4,183,355	11,075,748	302,424
Teshio	20	3,126,724	2,515,078	5,370,374	1,157,957

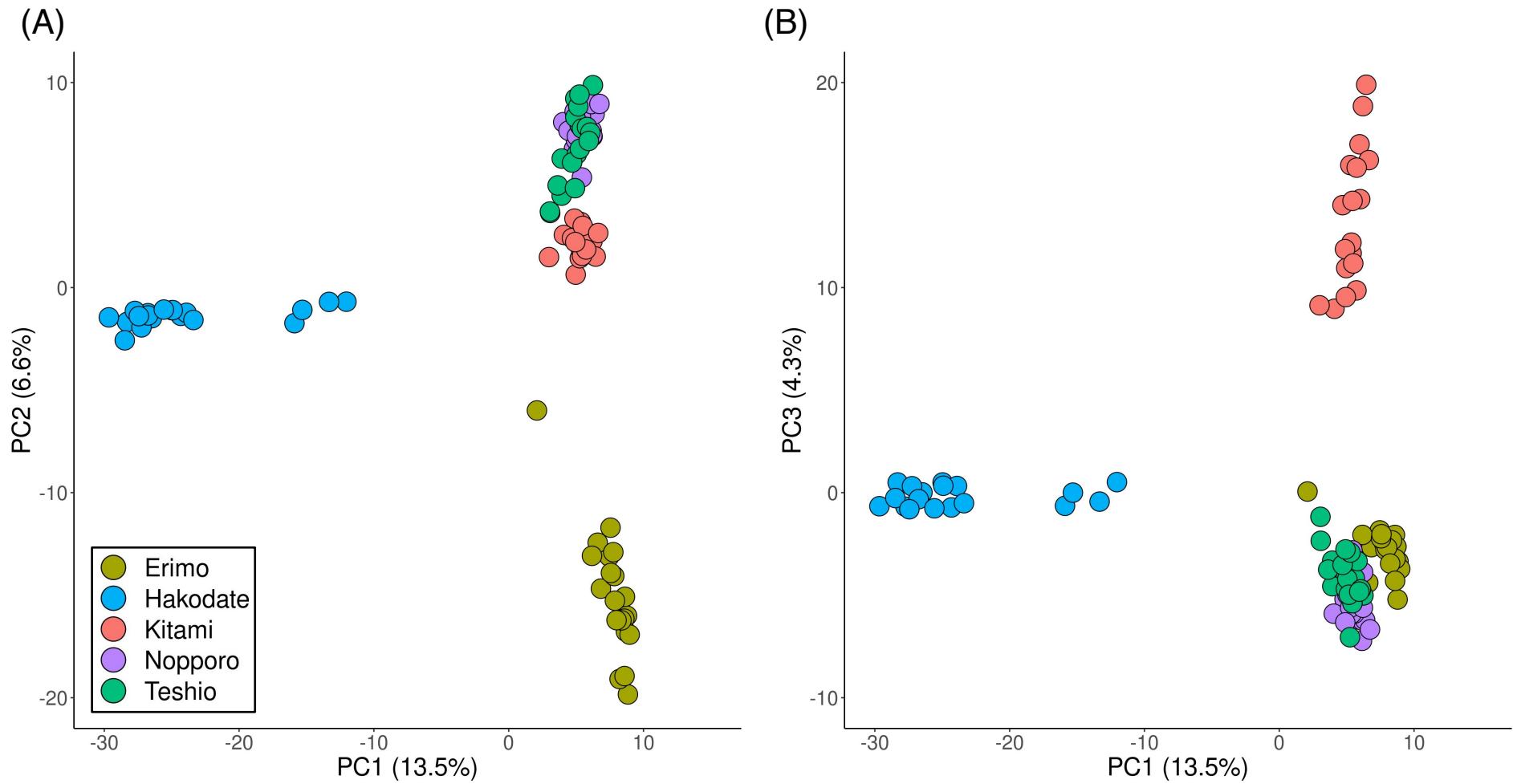
Result: SNP detection



Max Missing Count	No. of SNPs per locus	No. of all SNPs
0	432,459	717,676
35	859	6,962
25	547	4,510
15	193	2,237

検出したSNP座位には非常に欠損値が多かったため、96個体のうち、欠損値が25個体以下である547SNPsを以降の解析に用いた。

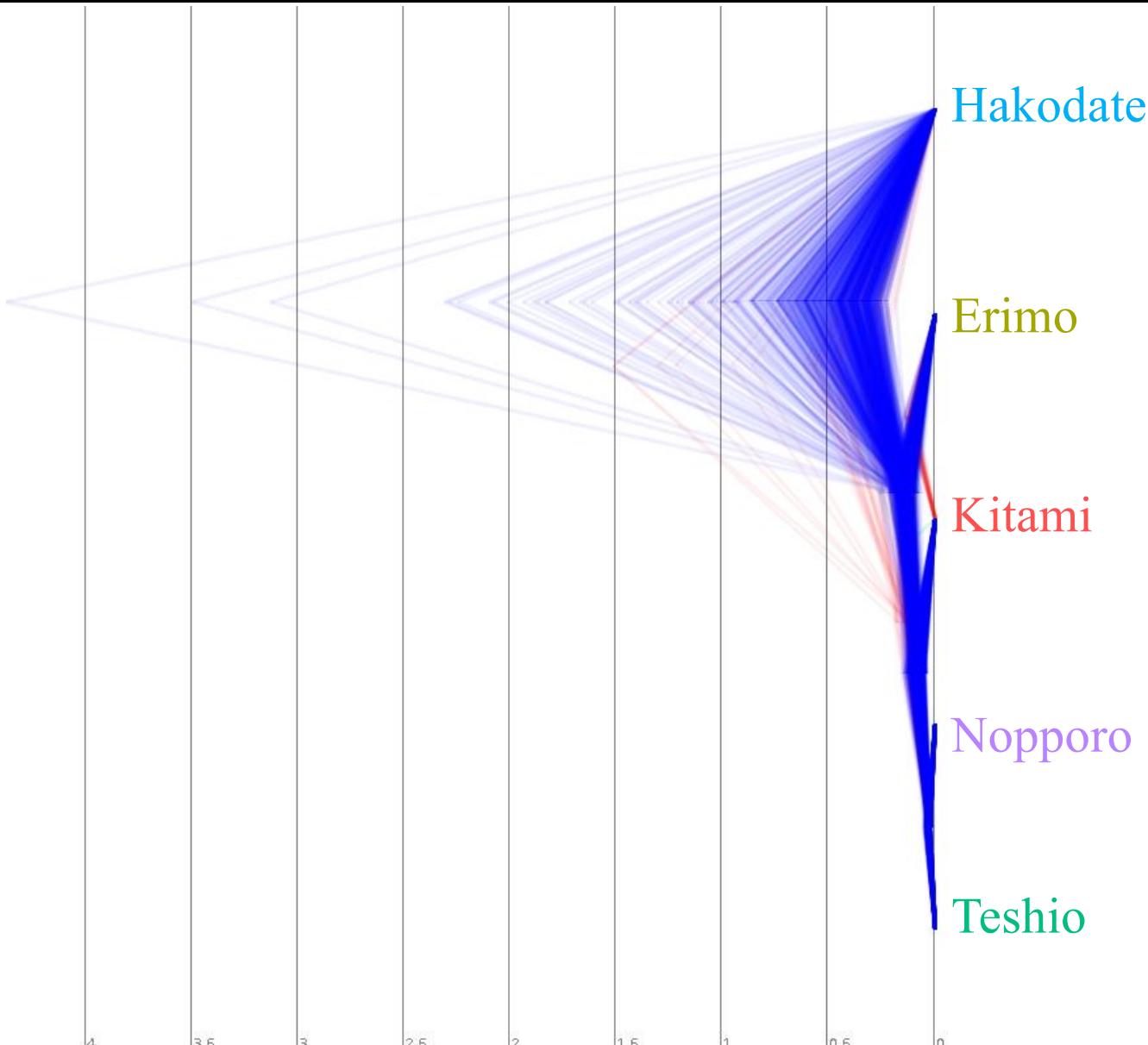
Result: PCA analysis



Matsunami et al. *under review*

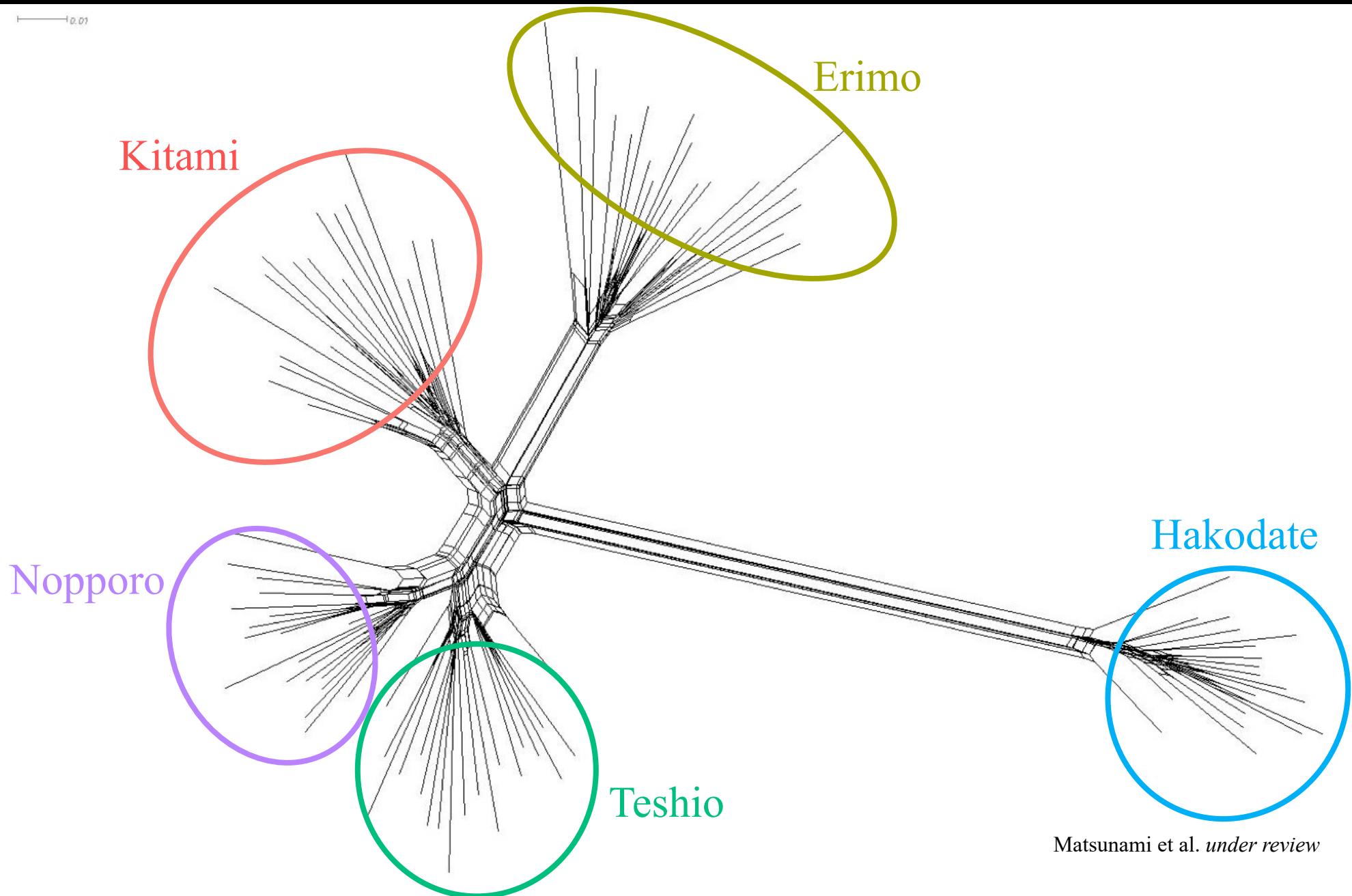
主成分分析の結果、野幌集団と天塩集団が遺伝的に非常に近縁であることが示唆された。

Result: SNAPP tree



Result: Phylogenetic network

10.01



Summary

- RAD-seq解析で、検出されたSNPの数が少なくても、それなりの解析(PCAなど)をすることは可能。
- シークエンシングエラーやmissing dataの取り扱いに注意することが大事。

第一回統合生物考古学若手研究集会: 参加者募集

開催概要

日程

2024/2/28 (水) 13:00-18:00 (18:00以降に同会場で懇親会)
2024/2/29 (木) 9:00-12:00

会場

北海道大学

開催形式

口頭・ポスター発表
対面形式（オンラインの予定はありません）



Location

北海道大学遠友学舎

会場へのアクセス



<https://bioarch-wakate.jimdofree.com/>

ご静聴ありがとうございました。

個別に質問があれば下記までお願いします。

E-mail: matsu@med.u-ryukyu.ac.jp

ハンズオンはgithubを使用します。

<https://github.com/mmatsunami/bioarch-2023>