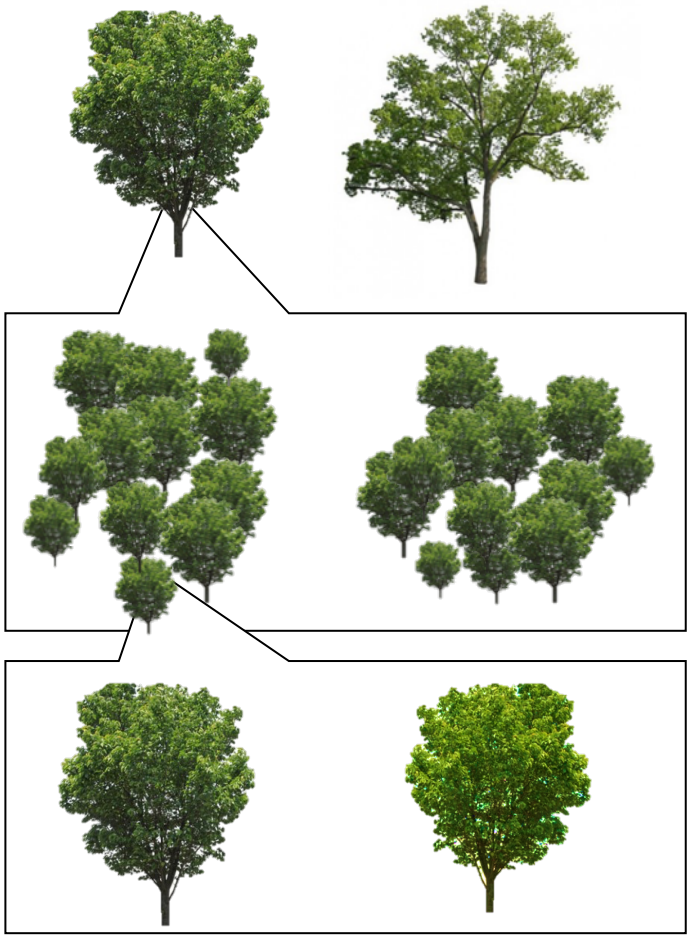
The 11th APBON Workshop
June 26, 2019

Introduction of next-generation biodiversity assessment using MIG-seq

陶山佳久　Yoshihisa Suyama
東北大学　Tohoku University, Japan

TOHOKU UNIVERSITY

Various levels of biological hierarchy to survey biodiversity



Among species

Among populations within a species

Among individuals within a population

Species A vs. B

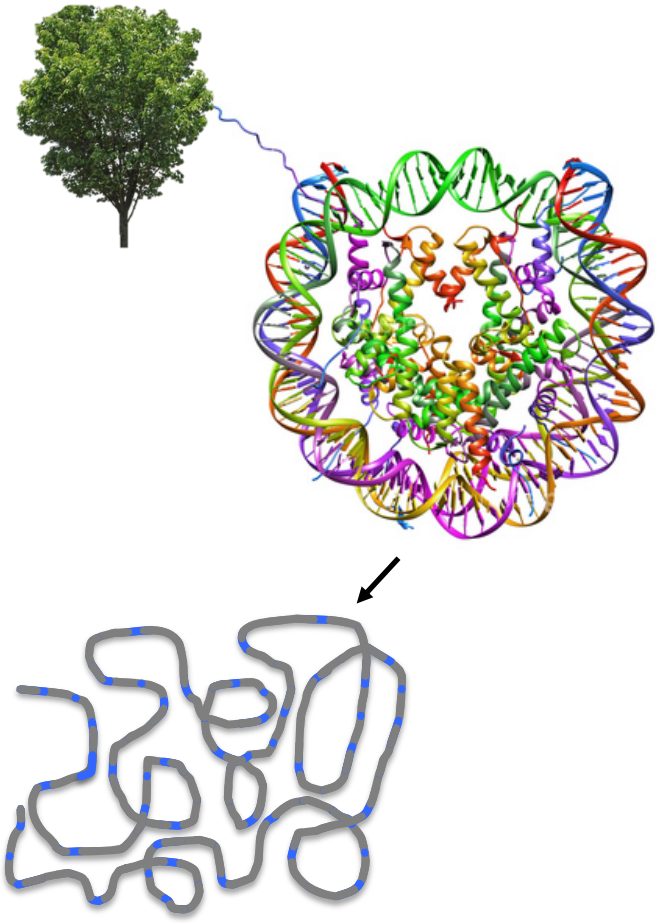
Populations A vs. B

Individuals A vs. B

→MIG-seq can be used for surveys of genetic differentiation of all these levels

Introduction

Whole genome sequence is not necessary to identify genetic differentiation



→MIG-seq is a genome wide sequencing method using reduced representation genomic library

Key Technique

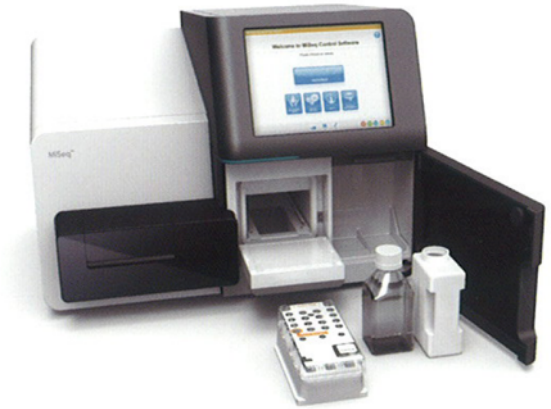
MIG-seq

Multiplexed ISSR Genotyping by sequencing

Suyama & Matsuki, 2015

Scientific Reports 5: 16963
doi: 10.1038/srep16963

Improved version will be available soon.



www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Received: 19 June 2015
Accepted: 22 October 2015
Published: 23 November 2015

MIG-seq: an effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform

Yoshihisa Suyama & Yu Matsuki

Restriction-enzyme (RE)-based next-generation sequencing methods have revolutionized marker-assisted genetic studies; however, the use of REs has limited their widespread adoption, especially in field samples with low-quality DNA and/or small quantities of DNA. Here, we developed a PCR-based procedure to construct reduced representation libraries without RE digestion steps, representing *de novo* single-nucleotide polymorphism discovery, and its genotyping using next-generation sequencing. Using multiplexed inter-simple sequence repeat (ISSR) primers, thousands of genome-wide regions were amplified effectively from a wide variety of genomes, without prior genetic information. We demonstrated: 1) Mendelian gametic segregation of the discovered variants; 2) reproducibility of genotyping by checking its applicability for individual identification; and 3) applicability in a wide variety of species by checking standard population genetic analysis. This approach, called multiplexed ISSR genotyping by sequencing, should be applicable to many marker-assisted genetic studies with a wide range of DNA qualities and quantities.

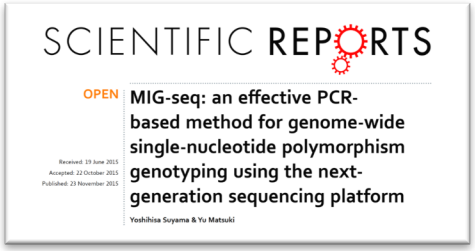
The recent development of next-generation sequencing (NGS) technology has allowed the effective discovery and genotyping of large numbers of genome-wide genetic markers[1]. However, many marker-assisted studies require more economical and efficient approaches, rather than the methods based on the high marker density produced by whole-genome sequencing. To optimize the cost and the amount of available information for these studies, several methods have been developed to construct reduced representation libraries (RRLs), to sample single-nucleotide polymorphisms (SNPs) from genome-wide regions, and to genotype them using NGS. Methods such as reduced representation shotgun[2] sequencing and restriction site-associated DNA (RAD)[3] markers were later adapted for NGS-based methods by the sequencing of RRLs[4] and RAD tags[5]. Many improved approaches have been developed in recent years, such as complexity reduction of polymorphic sequences[6], multiplexed shotgun sequencing[7], genotyping by sequencing (GBS)[8], 2-enzyme GBS[9], RAD genotyping using type IIB restriction enzymes[10], double digest RAD[11], and restriction fragment sequencing[12]. These methods have become widespread and allow marker-assisted genetic studies, such as ecological, evolutionary, phylogeographic, and genetic mapping studies, based on tens to hundreds of thousands of SNPs in hundreds of barcoded samples at the same time. However, a more simple, rapid and cost-efficient approach for smaller-scale studies is desired, especially in ecological laboratories.

Tohoku University, Kawatabi Field Science Center, Graduate School of Agricultural Science, 232-3 Yomogida, Naruko-onsen, Osaki, Miyagi 989-6711, Japan. Correspondence and requests for materials should be addressed to Y.S. (email: suyama@m.tohoku.ac.jp)

1. Brief introduction of MIG-seq

MIG-seq

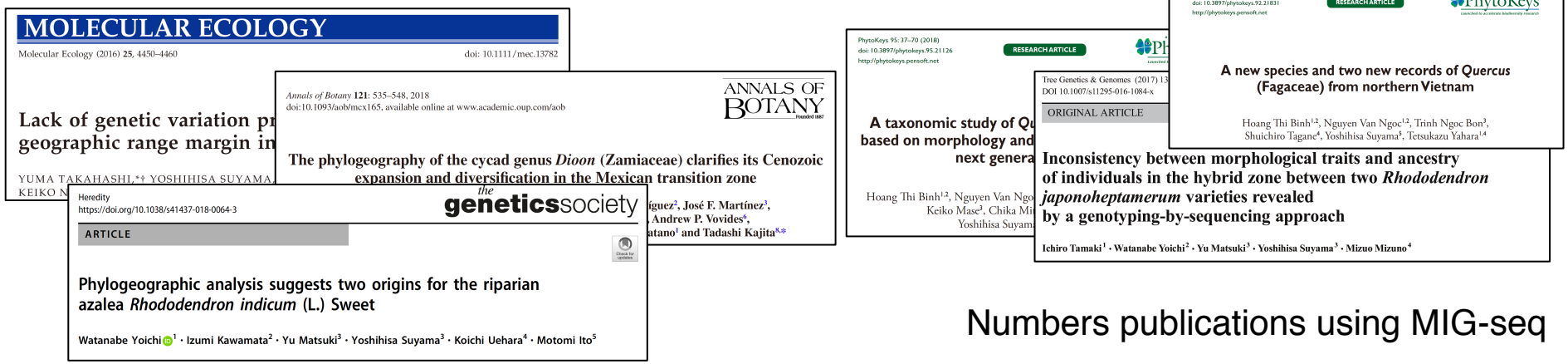
- Quick (Three days)
- Simple (Two PCRs & NGS run)
- Low cost (<10 USD/sample)
- >1000 of SNPs
- Low quality/quantity DNA
- A wide range of species


SCIENTIFIC REPORTS

OPEN

MIG-seq: an effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform

Yoshihisa Suyama & Yu Matsuki


MOLECULAR ECOLOGY

Molecular Ecology (2016) 25, 4450–4460 doi: 10.1111/mec.13782

Lack of genetic variation pr... geographic range margin in...

YUMA TAKAHASHI,*† YOSHIHISA SUYAMA... KEIKO N...

Annals of Botany 121: 535–548, 2018
doi:10.1093/aob/mcx165, available online at www.academic.oup.com/aob

ANNALS OF BOTANY

The phylogeography of the cycad genus *Dioon* (Zamiaceae) clarifies its Cenozoic expansion and diversification in the Mexican transition zone

...íguez[2], José F. Martínez[3], ...Andrew P. Vovides[6], ...atano[1] and Tadashi Kajita[8,*]

Heredity
https://doi.org/10.1038/s41437-018-0064-3

the geneticssociety

ARTICLE

Phylogeographic analysis suggests two origins for the riparian azalea *Rhododendron indicum* (L.) Sweet

Watanabe Yoichi[1] · Izumi Kawamata[2] · Yu Matsuki[3] · Yoshihisa Suyama[3] · Koichi Uehara[4] · Motomi Ito[5]

PhytoKeys 95: 37–70 (2018)
doi: 10.3897/phytokeys.95.21126
http://phytokeys.pensoft.net

RESEARCH ARTICLE

A taxonomic study of Q... based on morphology and ... next genera...

Hoang Thi Binh[1,2], Nguyen Van Ngo... Keiko Mase[3], Chika Mi... Yoshihisa Suyam...

Tree Genetics & Genomes (2017) 13
DOI 10.1007/s11295-016-1084-x

ORIGINAL ARTICLE

Inconsistency between morphological traits and ancestry of individuals in the hybrid zone between two *Rhododendron japonoheptamerum* varieties revealed by a genotyping-by-sequencing approach

Ichiro Tamaki[1] · Watanabe Yoichi[2] · Yu Matsuki[3] · Yoshihisa Suyama[3] · Mizuo Mizuno[4]

PhytoKeys 92: 1–15 (2018)
doi: 10.3897/phytokeys.92.21831
http://phytokeys.pensoft.net

RESEARCH ARTICLE

PhytoKeys

A new species and two new records of *Quercus* (Fagaceae) from northern Vietnam

Hoang Thi Binh[1,2], Nguyen Van Ngoc[1,2], Trinh Ngoc Bon[3], Shuichiro Tagane[4], Yoshihisa Suyama[4], Tetsukazu Yahara[1,4]

Numbers publications using MIG-seq

1. Brief introduction of MIG-seq

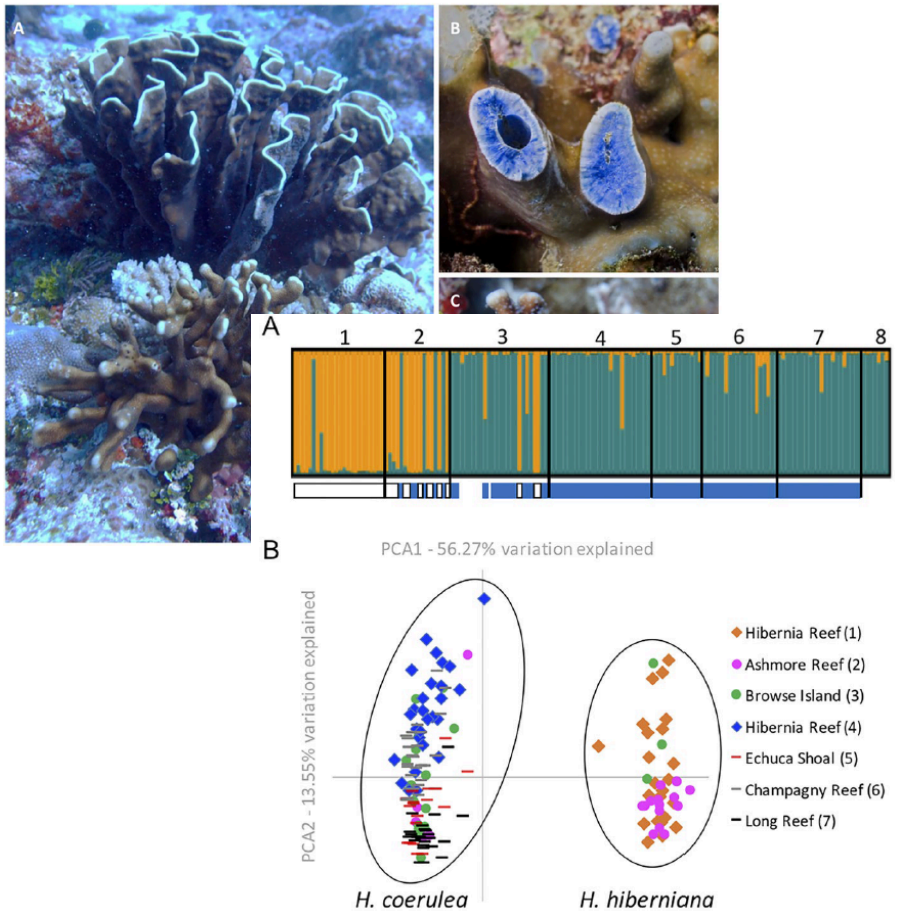
SCIENTIFIC REPORTS

OPEN

Integrated evidence reveals a new species in the ancient blue coral genus *Heliopora* (Octocorallia)

Received: 16 April 2018
Accepted: 17 September 2018
Published online: 26 October 2018

Zoe T. Richards[1,2], Nina Yasuda[3], Taisei Kikuchi[4], Taryn Foster[5], Chika Mitsuyuki[6], Michael Stat[2,7], Yoshihisa Suyama[6] & Nerida G. Wilson[1,8]



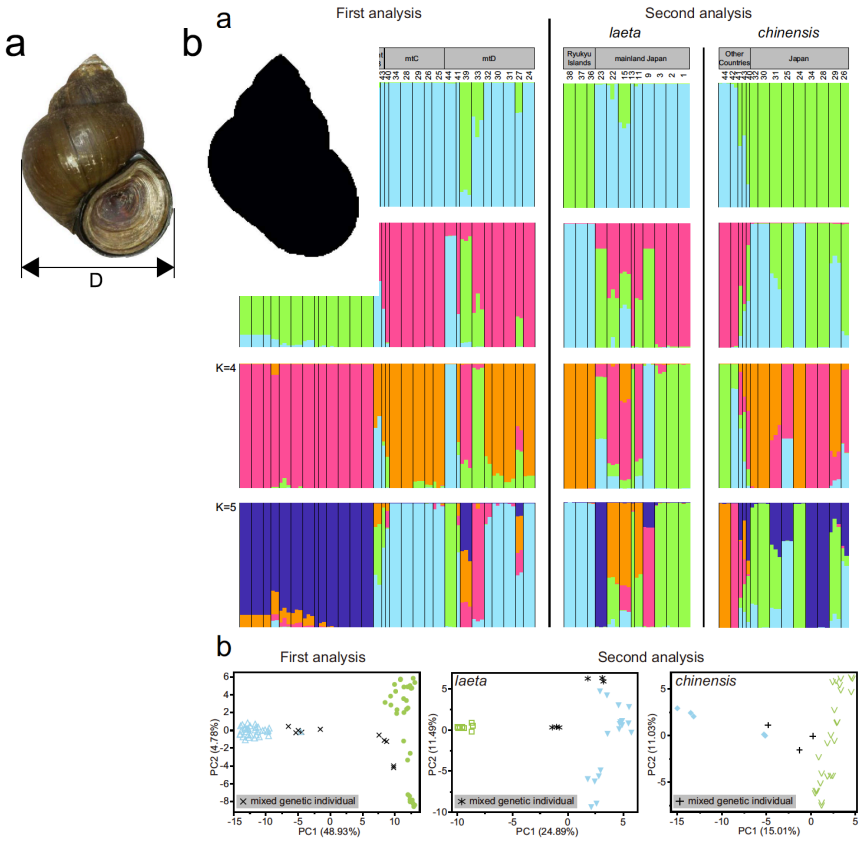
SCIENTIFIC REPORTS

OPEN

Enigmatic incongruence between mtDNA and nDNA revealed by multi-locus phylogenomic analyses in freshwater snails

Received: 28 August 2018
Accepted: 5 April 2019
Published online: 17 April 2019

Takahiro Hirano[1], Takumi Saito[2], Yoshihiro Tsunamoto[3], Joichiro Koseki[2], Bin Ye[2,4], Van Tu Do[5], Osamu Miura[6], Yoshihisa Suyama[3] & Satoshi Chiba[2,7]

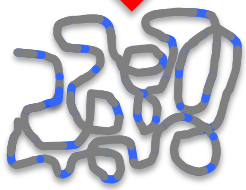


1. Brief introduction of MIG-seq

Leaf Materials

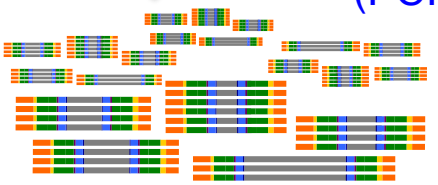
DNA extraction

DNA



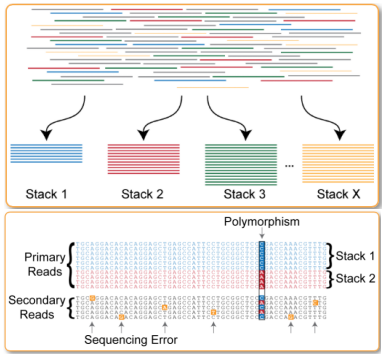
Library Construction (PCR)

NGS Library



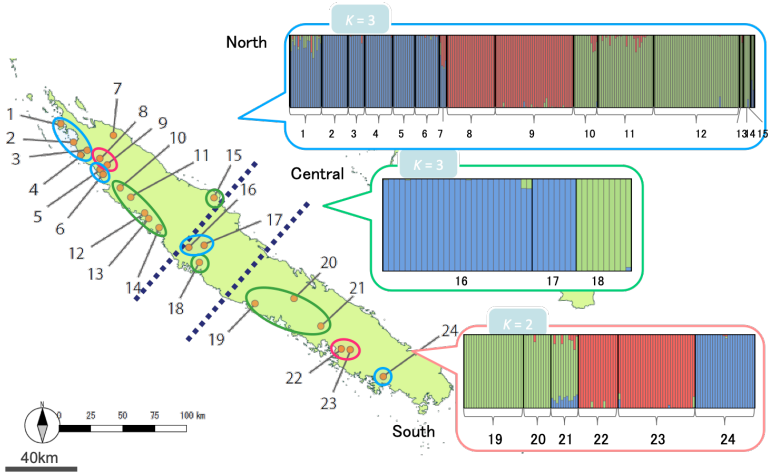
NGS Run

>Millions of Sequences

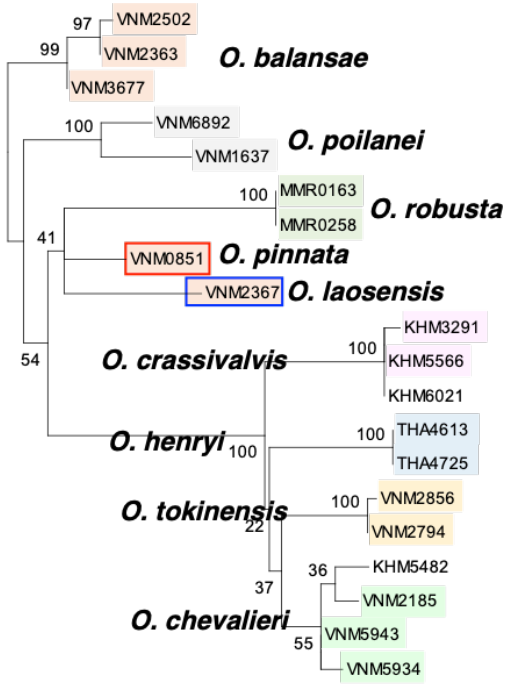


Data Analysis

- Quick (Three days)
- Simple (PCR-based)
- Economical (~10 USD / sample)



Population Genetics and Phylogeography



Phylogenetics

MIG-seq: Multiplexed ISSR Genotyping by sequencing

PCR-based procedure for constructing reduced representation libraries, involving *de novo* SNP discovering, and their genotyping using NGS

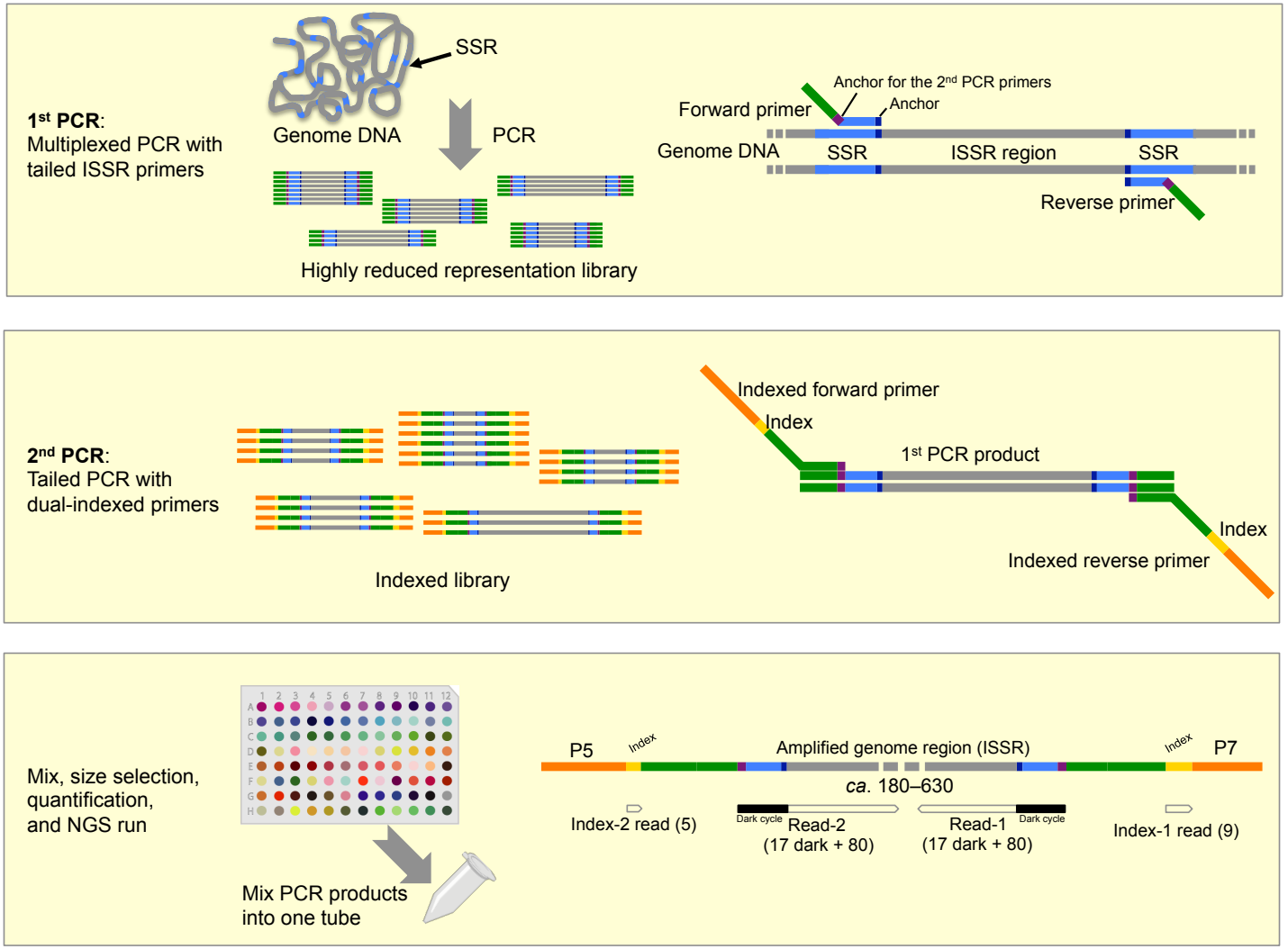
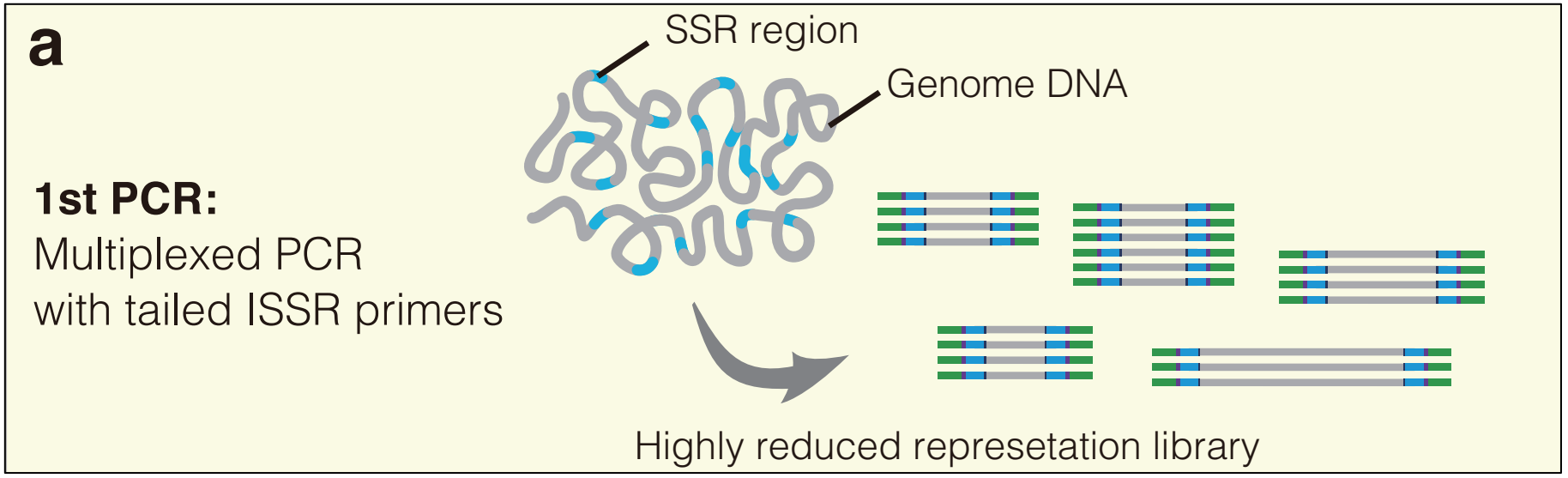
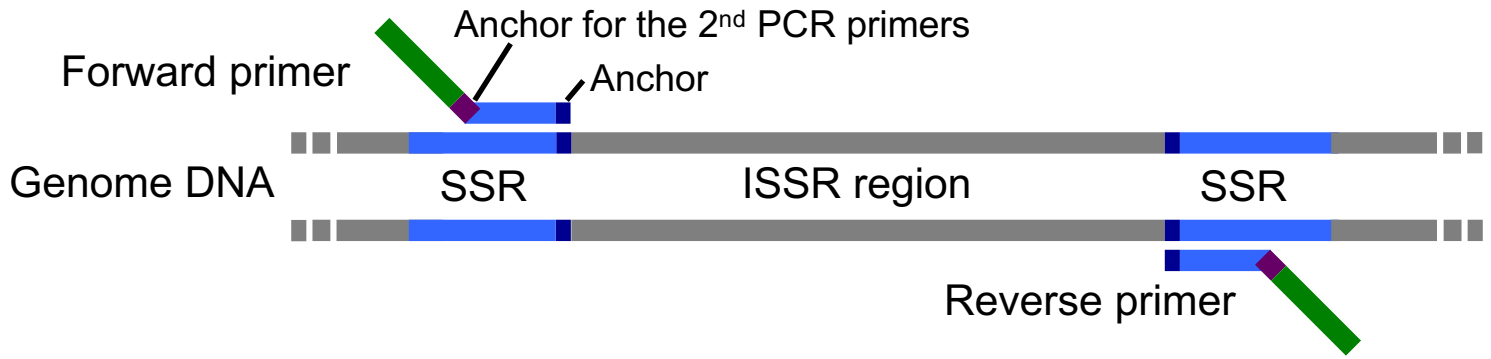


Fig. Three steps of MIG-seq procedure

Methods: Step #1

1st PCR: An effective PCR amplification of thousands of genome-wide regions from a genome





Multiple non-repetitive regions from various inter-simple-sequence repeat (ISSR) are amplified from genomic DNA by multiplexed PCR with universal tailed-ISSR primers.

Methods: Step #1

1st PCR

Tailed ISSR primers set

Table 1. Sequences of MIG-seq primer set-1 for the 1st PCR

Name	Sequences (5′–3′)
Forward primers:	(Tail + **anchor: CTG**) + SSR + **anchor**
$(ACT)_4$TG-f	CGCTCTTCCGATCT**CTG**ACTACTACTACT**TG**
$(CTA)_4$TG-f	CGCTCTTCCGATCT**CTG**CTACTACTACTA**TG**
$(TTG)_4$AC-f	CGCTCTTCCGATCT**CTG**TTGTTGTTGTTG**AC**
$(GTT)_4$CC-f	CGCTCTTCCGATCT**CTG**GTTGTTGTTGTT**CC**
$(GTT)_4$TC-f	CGCTCTTCCGATCT**CTG**GTTGTTGTTGTT**TC**
$(GTG)_4$AC-f	CGCTCTTCCGATCT**CTG**GTGGTGGTGGTG**AC**
$(GT)_6$TC-f	CGCTCTTCCGATCT**CTG**GTGTGTGTGTGT**TC**
$(TG)_6$AC-f	CGCTCTTCCGATCT**CTG**TGTGTGTGTGTG**AC**
Reverse primers:	(Tail + **anchor: GAC**) + SSR + **anchor**
$(ACT)_4$TG-r	TGCTCTTCCGATCT**GAC**ACTACTACTACT**TG**
$(CTA)_4$TG-r	TGCTCTTCCGATCT**GAC**CTACTACTACTA**TG**
$(TTG)_4$AC-r	TGCTCTTCCGATCT**GAC**TTGTTGTTGTTG**AC**
$(GTT)_4$CC-r	TGCTCTTCCGATCT**GAC**GTTGTTGTTGTT**CC**
$(GTT)_4$TC-r	TGCTCTTCCGATCT**GAC**GTTGTTGTTGTT**TC**
$(GTG)_4$AC-r	TGCTCTTCCGATCT**GAC**GTGGTGGTGGTG**AC**
$(GT)_6$TC-r	TGCTCTTCCGATCT**GAC**GTGTGTGTGTGT**TC**
$(TG)_6$AC-r	TGCTCTTCCGATCT**GAC**TGTGTGTGTGTG**AC**

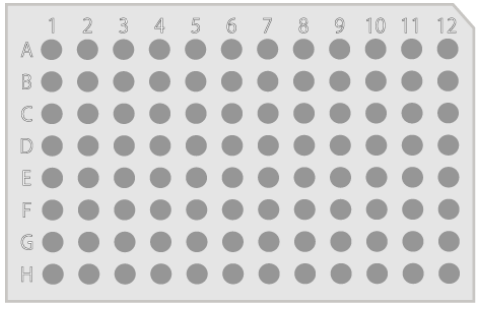
Underlined and boldface nucleotides denote tail and anchor sequences, respectively. The difference between forward and reverse primer sets is only in their tail sequences.

Methods: Step #1
1st PCR

Forward primers:	(Tail + **anchor: CTG**) + SSR + **anchor**
$(ACT)_4TG$-f	CGCTCTTCCGATCT**CTG**ACTACTACTACT**TG**
$(CTA)_4TG$-f	CGCTCTTCCGATCT**CTG**CTACTACTACTA**TG**
$(TTG)_4AC$-f	CGCTCTTCCGATCT**CTG**TTGTTGTTGTTG**AC**
$(GTT)_4CC$-f	CGCTCTTCCGATCT**CTG**GTTGTTGTTGTT**CC**
$(GTT)_4TC$-f	CGCTCTTCCGATCT**CTG**GTTGTTGTTGTT**TC**
$(GTG)_4AC$-f	CGCTCTTCCGATCT**CTG**GTGGTGGTGGTG**AC**
$(GT)_6TC$-f	CGCTCTTCCGATCT**CTG**GTGTGTGTGTGT**TC**
$(TG)_6AC$-f	CGCTCTTCCGATCT**CTG**TGTGTGTGTGTG**AC**
Reverse primers:	(Tail + **anchor: GAC**) + SSR + **anchor**
$(ACT)_4TG$-r	TGCTCTTCCGATCT**GAC**ACTACTACTACT**TG**
$(CTA)_4TG$-r	TGCTCTTCCGATCT**GAC**CTACTACTACTA**TG**
$(TTG)_4AC$-r	TGCTCTTCCGATCT**GAC**TTGTTGTTGTTG**AC**
$(GTT)_4CC$-r	TGCTCTTCCGATCT**GAC**GTTGTTGTTGTT**CC**
$(GTT)_4TC$-r	TGCTCTTCCGATCT**GAC**GTTGTTGTTGTT**TC**
$(GTG)_4AC$-r	TGCTCTTCCGATCT**GAC**GTGGTGGTGGTG**AC**
$(GT)_6TC$-r	TGCTCTTCCGATCT**GAC**GTGTGTGTGTGT**TC**
$(TG)_6AC$-r	TGCTCTTCCGATCT**GAC**TGTGTGTGTGTG**AC**

TaKaRa

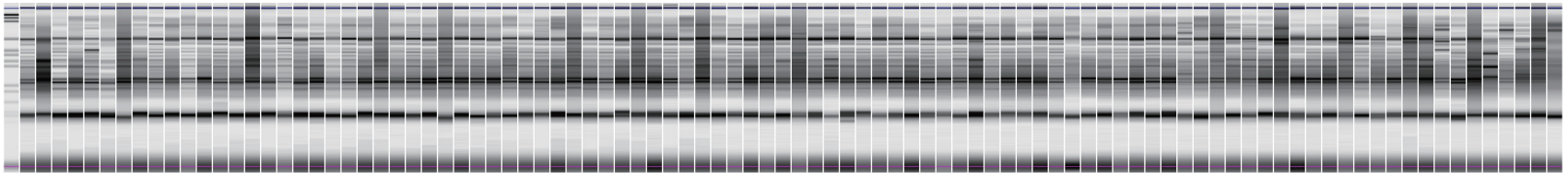
Multiplex PCR Assay Kit Ver. 2



Template DNA	1.0 µL
[Mixture]	
2x Multiplex PCR Buffer	10.0 µL
1st PCR primers	3.2 µL each @20 µM
Multiplex PCR Enzyme mix	0.1 µL
Water	5.7 µL
	20.0 µL

94°C	1 min	
94°C	30 sec	|
38°C *	1 min	| 18–25 cycles
72°C	1 min	|
72°C	10 min	

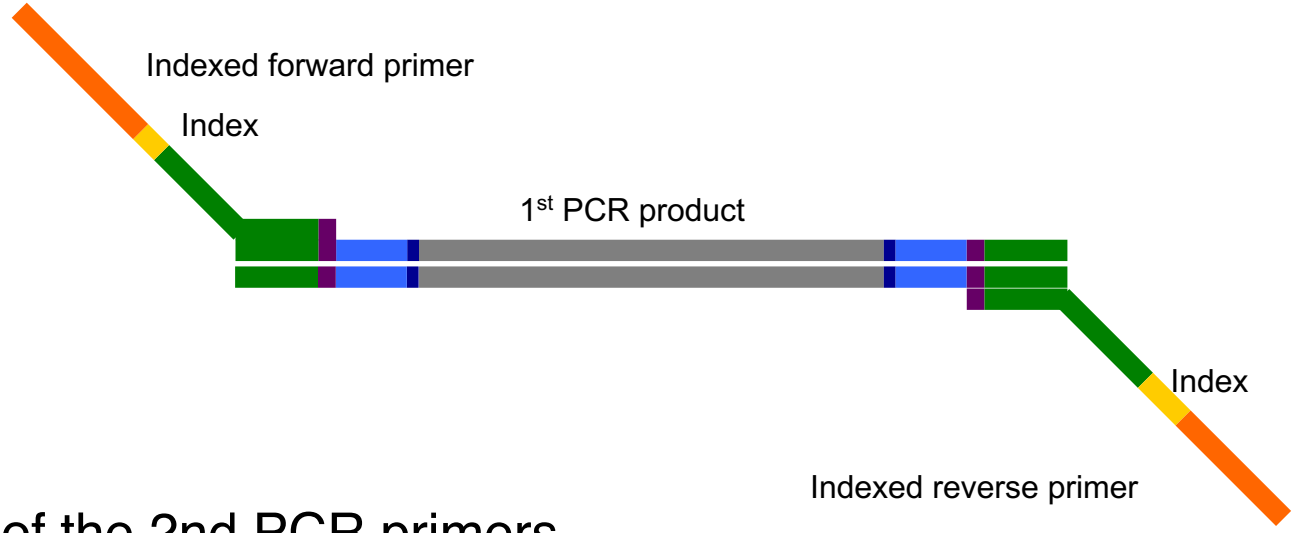
*changed from published version



Methods: Step #2

2nd PCR (Tailed PCR)

This step enables the addition of complementary sequences for the binding sites of Illumina sequencing flow cell and index (barcode) for each sample to the 1st PCR products using indexed primers.



New version of the 2nd PCR primers

Index (5 bases)

Forward (index) **AATGATACGGCGACCACCGAGATCTACACxxxxxACACTCTTTCCCTACACGACGCTCTTCCGATCTCTG**

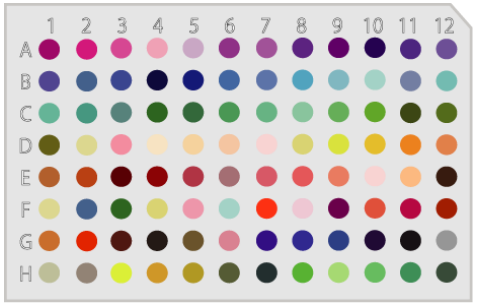
Reverse (index) **CAAGCAGAAGACGGCATACGAGATxxxxxxxxxGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGAC**

Index (9 bases)

Methods: Step #2
2nd PCR (Tailed PCR)

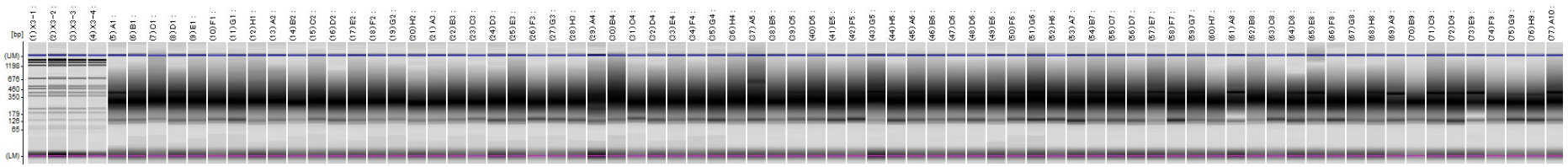
TaKaRa

PrimeSTAR® GXL DNA Polymerase



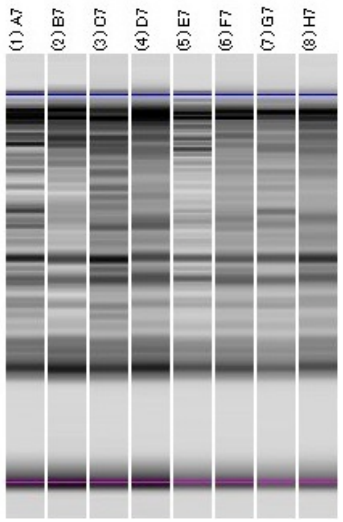
Diluted 1st PCR product*1	2.5 µL
5x PrimeSTAR GXL Buffer	2.4 µL
dNTP mixture	0.96 µL
2nd PCR forward primer	1.2 µL @2 µM
2nd PCR reverse primer*2	1.2 µL @2 µM
PrimerSTAR GXL polymerase	0.24 µL
Water	3.5 µL
	12.0 µL

98°C	10 sec	|
54°C	15 sec	| 12 cycles
68°C	1 min	|



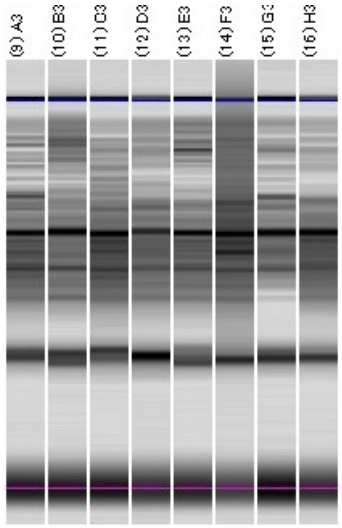
Methods: Steps #3-1: Mix, Size selection, and Quantification

1st PCR
(Multiplexed Tailed PCR)



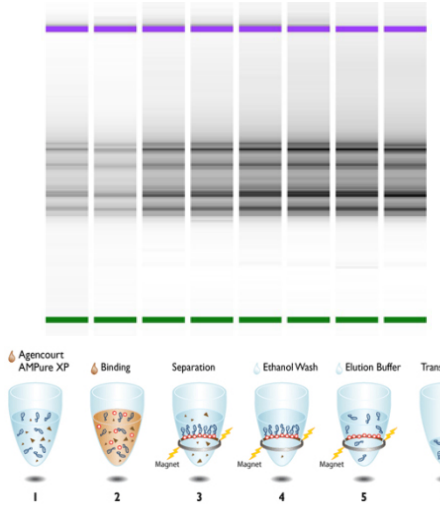
→

2nd PCR
(Tailed PCR + Index)



→

Pool,
Size selection
(350–800 bp)



(AMPure XP, Beckman Coulter)

Construction of the MIG-seq library

Methods: Step #3-2
NGS Run

MiSeq Reagent kit v3 (150 cycle, Illumina)

MiSeq Sequencer (Illumina)

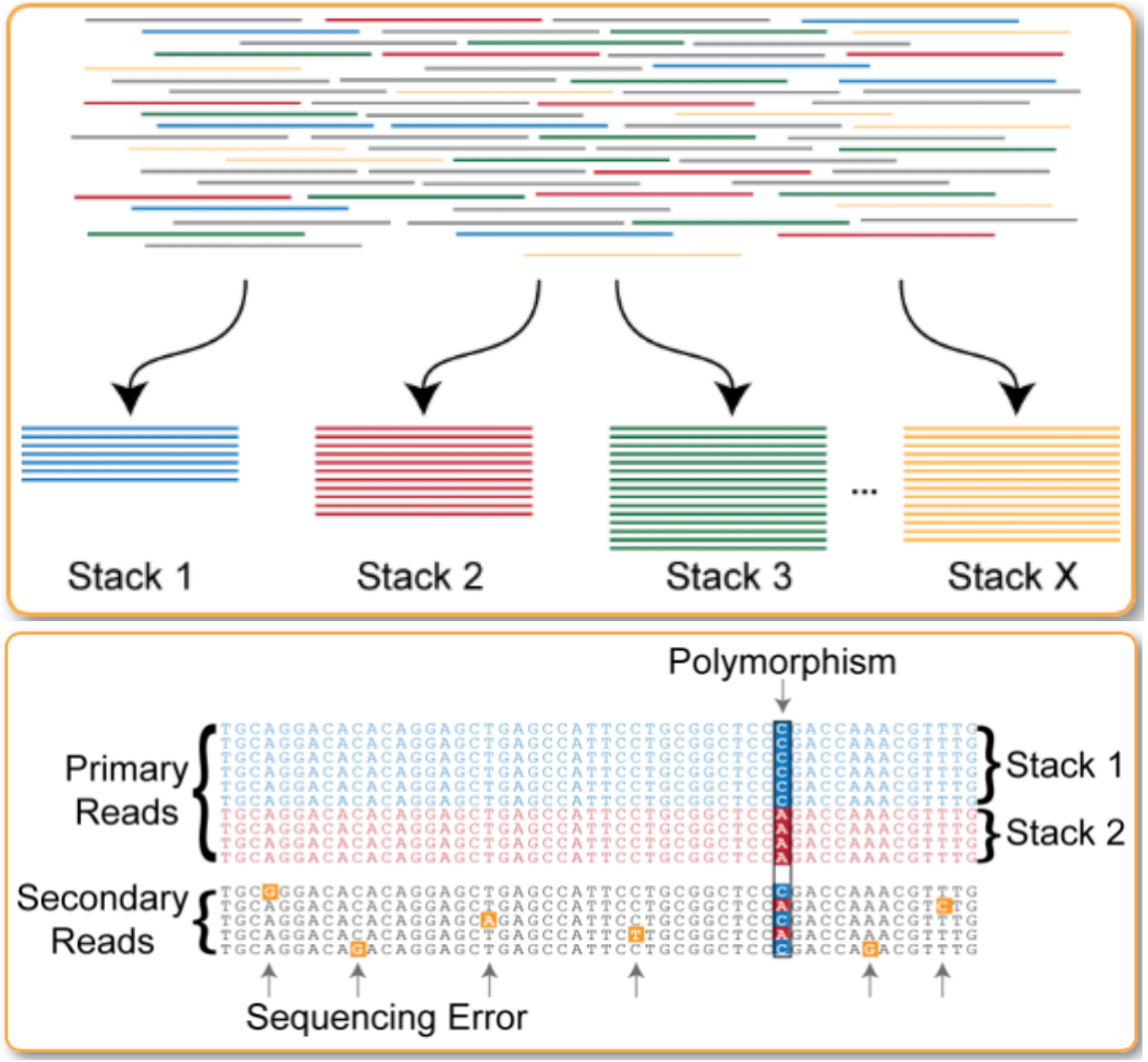
Amount of data from 96–192 samples of a standard species using MiSeq

No. of raw reads:	ca. 50～60 million
No. of reads/sample:	~600,000 reads
No. of SNP markers:	~5000 SNPs or more

Methods: Data analysis

Stacks

pyRAD, and so on.



Data analysis with Stacks（Catchen et al 2011）

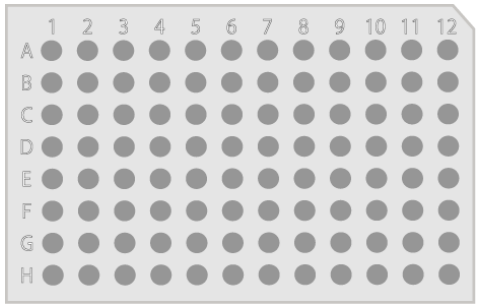
Time and cost

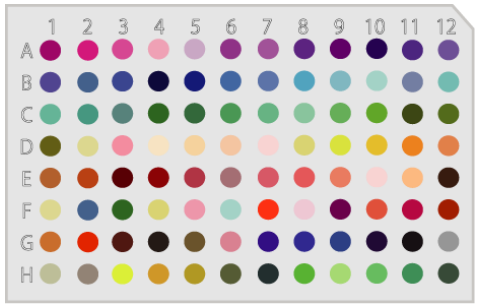
DNA extraction
（PCR quality）

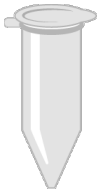
1st PCR
96–192 samples
(even more)

2nd PCR
adding index for
each sample

Pool,
Size select,
Quantify







~384 samples (or more)
(without genome info.)

Time： 3 days
Library construction: 1 day
NGS run: 1 day
Data analysis: 1 day

Cost：
~1,400 US$
(4~ US$/sample)

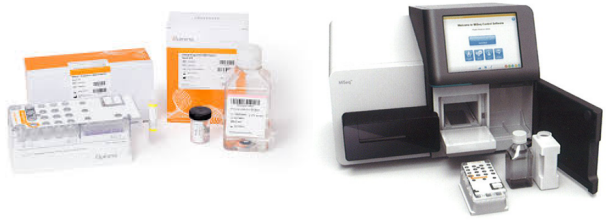
Calculation is based on
run with MiSeq platform
and
MiSeq Reagent kit v3
(150 cycle, Illumina)

Multiplexed DNA Barcoding

A simple and economical protocol to detect sequences of ITS and several cpDNA regions (*psb*A-*trn*H, *rbc*L, *trn*L^{UAA}intron and *mat*K) together using one run of NGS.

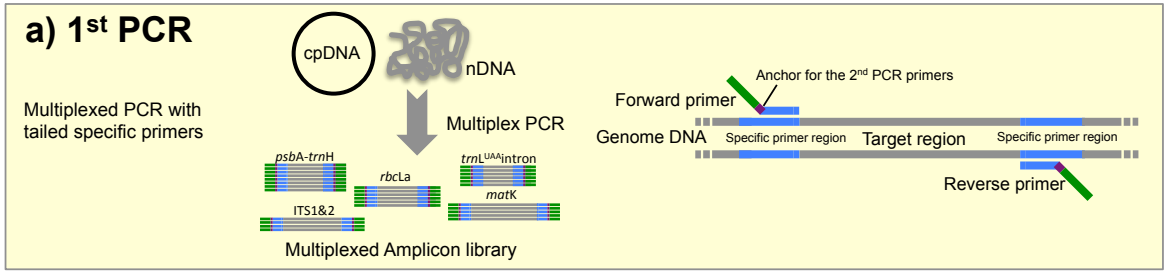
Primers for multiplexed DNA barcoding

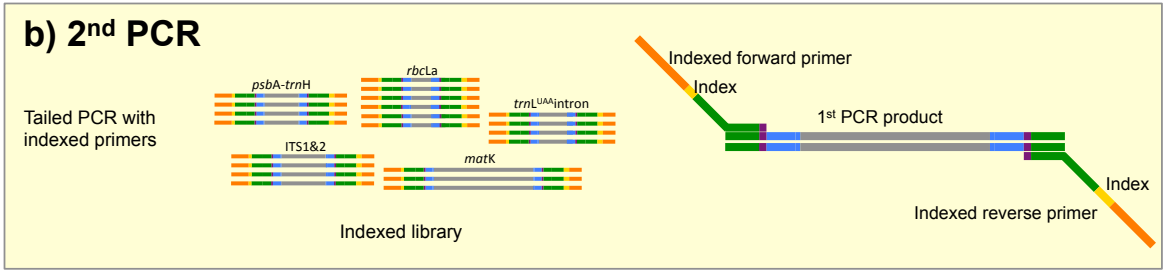
Region	Primer name	Sequence (5'-3')	Ref.
*psb*A-*trn*H	psbAF_f-tail	f-tail + GTTATGCATGAACGTAATGCTC	1
	trnHR_r-tail	r-tail + CGCGCATGGTGGATTCACAATCC	
*mat*K	KIM 3F_f-tail	f-tail + CGTACAGTACTTTTGTGTTTACGAG	2
	KIM 1R_r-tail	r-tail + ACCCAGTCCATCTGGAAATCTTGGTTC	
*rbc*La	rbcLa_F_f-tail	f-tail + ATGTCACCACAAACAGAGACTAAAGC	3
	rbcLa_R_r-tail	r-tail + GTAAAATCAAGTCCACCRCG	
ITS1&2	ITS-u1_f-tail	f-tail + GGAAGKARAAGTCGTAACAAGG	4
	ITS-u4_r-tail	r-tail + RGTTTCTTTTCCTCCGCTTA	
*trn*L^{UAA}intron	trnL^{UAA}c_f-tail	f-tail + CGAAATCGGTAGACGCTACG	5
	trnL^{UAA}d_r-tail	r-tail + GGGGATAGAGGGACTTGAAC	

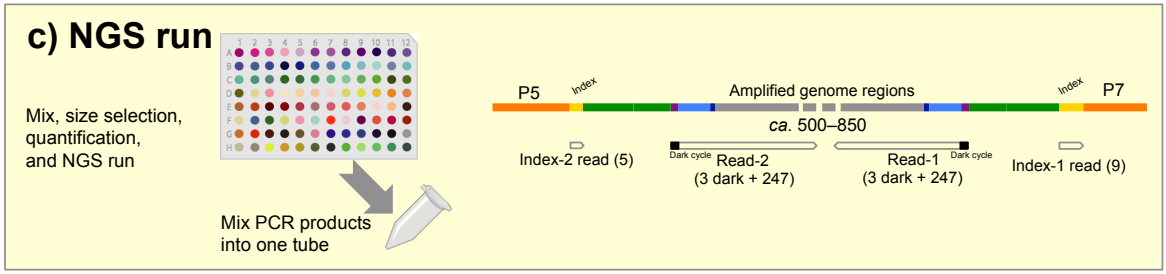


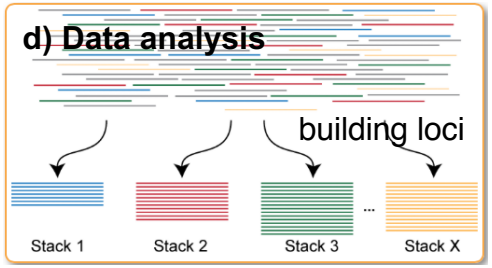
Illumina MiSeq Reagent Nano Kit v2 (500 cycles)
=*ca.* 500 USD for 384 samples
(< 2 USD/sample)

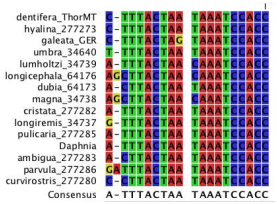
~250 bp from both ends of each region



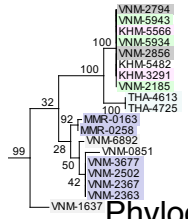








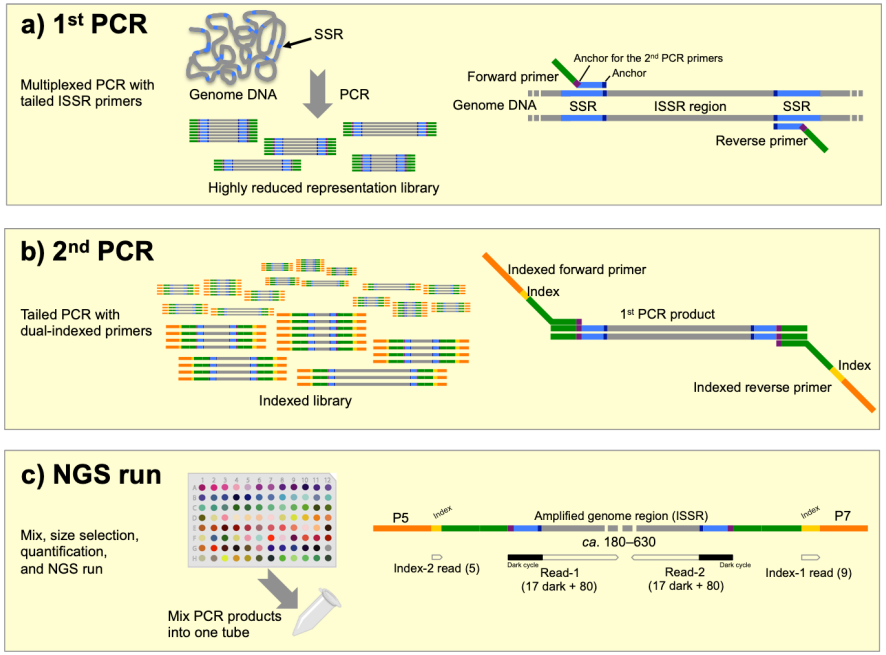
Alignment

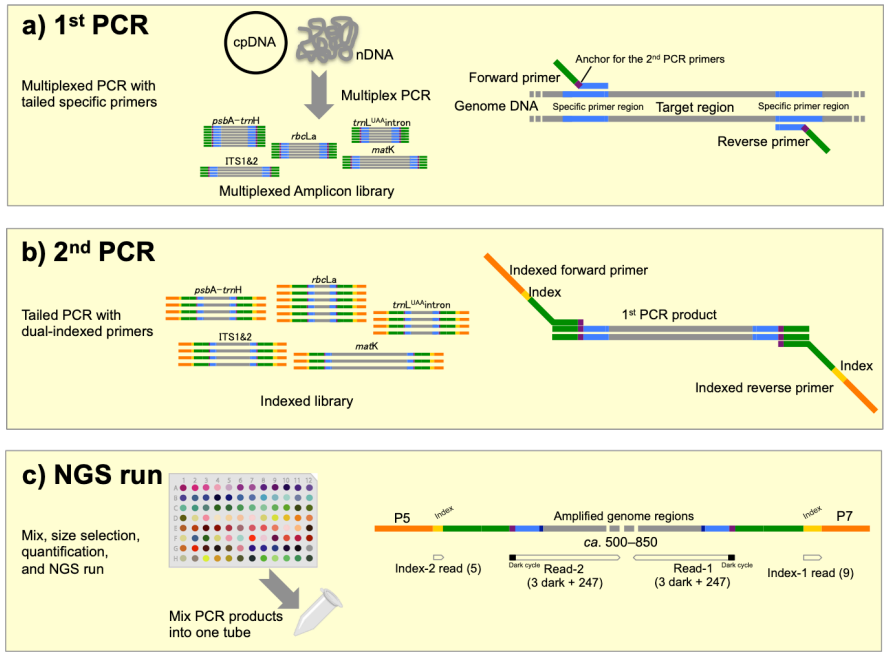


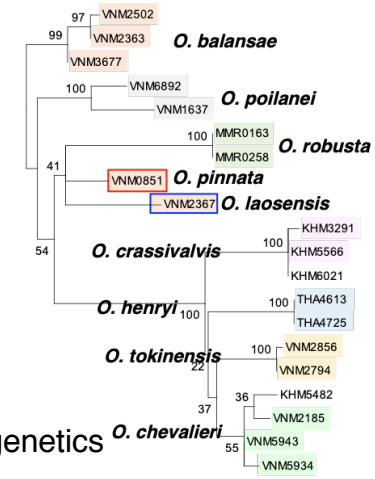
Phylogenetics

MIG-seq and Multiplexed DNA barcoding

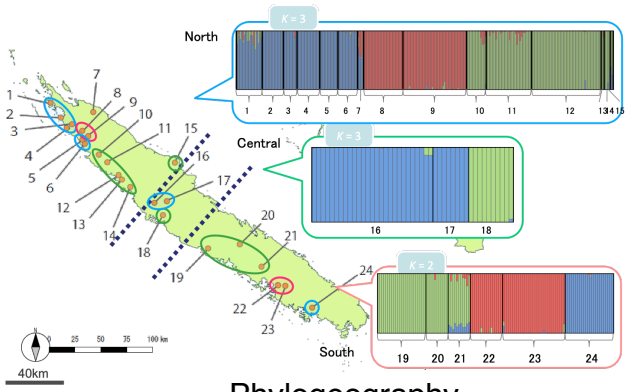
An efficient combination for the next-generation biodiversity assessment



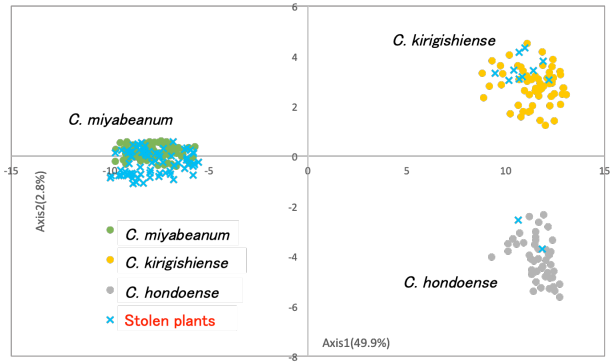




Phylogenetics



Phylogeography



Conservation Genetics

MIG-seq and Multiplexed DNA Barcoding

New tools for biodiversity assessment

Three independent genetic data based on
1) Genome-wide SNPs
2) cpDNA genes (or mtDNA, ex. CO1)
3) ITS, etc. can be easily detected

MIG-seq can cover: family, genus, species, population, individual
Multi-barcoding: family, genus, species, population, individual

Advantages

- Applicable to a wide range of DNA quality/quantity
- Quick (three days + one–three days)
- Simple (two PCR steps and NGS run)
- Economical (< 2000 USD / >384 samples)

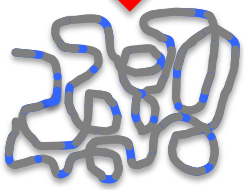
MIG-seq Analysis Package

Supported by grants from the Project of the NARO Bio-oriented Technology Research Advancement Institution (Research program on development of innovative technology)

Materials

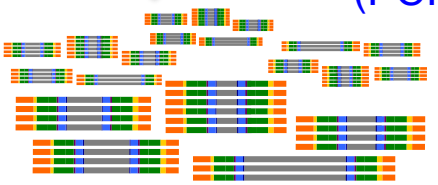
DNA extraction

DNA



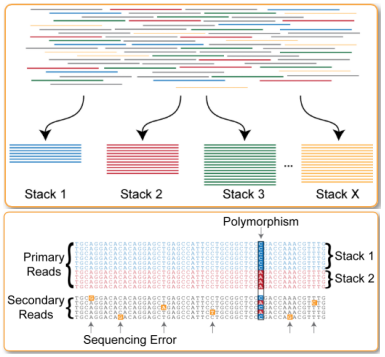
Library Construction (PCR)

NGS Library



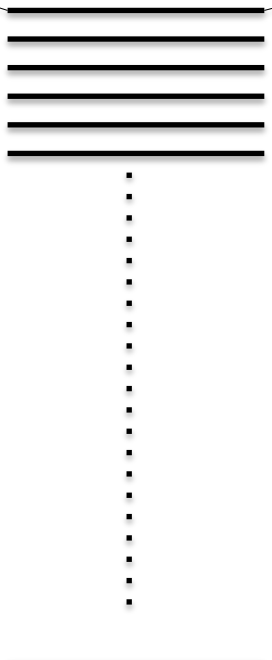
NGS Run

>Millions of Sequences



MIG-seq Data

ATGAAGGTAAATGTGAAGATATTGGTGAATAATGGAGTGAAAATAGGGACCATTTTAGGCCCAACAAACACACCAGAGAT



Identification

MIG-seq Data Analysis System

MIG-seq Database

MIG-seq Species Identification System

Summary: Features of MIG-seq

TOHOKU UNIVERSITY

Advantages

- Applicable to a wide range of DNA quality/quantity
- Quick (three days)
- Simple (two PCR steps and NGS run)
- Economical (~10 US$/sample)

Disadvantages

- Inapplicable to >10000 markers
- SNPs from limited regions

Recommended studies

- Population genetics, phylogeography and phylogeny
- Identification of clones, hybrids, breeding varieties, and species
- Next-generation biodiversity assessment

Acknowledgements

Methods

Yu Matsuki
Ayumi Matsuo
Shun Hirota
Chica Mitsuyuki
Yoshihiro Tsunamoto
Mitsuhiko Sato

New Caledonia

Yuji Isagi
Kotomi Fujita
Gildas Gâteblé
Rimpei Nagaoka

SE Asia

Tetsukazu Yahara
Shuichiro Tagane
Hironori Toyama
Akiyo Naiki
Etsuko Moritsuka

Funding

Environment Research and Technology Development Fund (4-1605, 4-1902) of the Environmental Restoration and Conservation Agency of Japan

JSPS KAKENHI Grant Numbers JP16H02553, JP18H04011, JP1802469,

NARO Bio-oriented Technology Research Advancement Institution (Research program on development of innovative technology), Grant Number 30031C

独立行政法人
環境再生保全機構
ERCA
Environmental Restoration and Conservation Agency

科研費
KAKENHI

農研機構
NARO
国立研究開発法人 農業・食品産業技術総合研究機構

TOHOKU
UNIVERSITY