

분자마커 개발 실습

(주)바이오투
전임연구원 강윤주

목차

Part I. SNP 변이 선별 및 주변 서열 확보	Part II. Primer 디자인 및 결과 확인
1. VCF 파일에서 변이 확인	1. 'Primer3'를 통한 PCR primer design
2. 'IGV'를 통한 변이 검증 (IGV 사용법)	2. 'Bowtie'를 통한 PCR primer 검증
3. Polymorphism 확인	3. Unique PCR primer 선별
4. 주변 서열 확보	4. 결과 확인

분석과정 필수 프로그램

□ IGV (Integrative Genomics Viewer)

- 통합적인 유전체 데이터셋을 시각화해주는 그래픽 기반 프로그램
- 다양한 유전체 관련 정보를 여러가지 트랙을 통하여 보여줌

□ Primer3

- PCR primer 를 디자인하는데 사용하는 프로그램

□ Bowtie

- Short read aligner
- 빠르고 효율적으로 유전체서열에 read align 하는 프로그램

□ VI / VIM

- 리눅스에서 활용 가능한 텍스트 편집기

Part I. SNP 변이 선별 및 주변 서열 확보

1. VCF 파일에서 변이 확인

□ 작성된 VCF 파일을 통해 SNP 변이를 확인

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	MixR	MixS	P1	P2
##source=CombineGVCFs												
##source=SelectVariants												
##source=VariantFiltration												
##bcftools_viewVersion=1.17+htslib-1.17												
##bcftools_viewCommand=view -i FILTER="PASS" combinedflt_snps.vcf; Date=Wed Jul 5 14:01:03 2023												
chr01	80388	.	C	T	40.18	PASS	AC=2;AF=	GT:AD:DP:	./.:0,0:0:0,0,0,0	1/1:0,2:2:6:49,6	./.:0,0:0:0,0,0,0	./.:0,0:0:0,0,0,0
chr01	159218	.	G	A	60.04	PASS	AC=2;AF=	GT:AD:DP:	./.:0,0:0:0:0,0,0,0	0/0:1,0:1:3:0,0,0	1 1:0,2:2:6:1 1:	./.:1,0:1:0:0,0,0,0
chr01	159225	.	G	A	36.08	PASS	AC=2;AF=	GT:AD:DP:	./.:0,0:0:0:0,0,0,0	0/0:1,0:1:3:0,0,0	1 1:0,1:1:3:1 1:	0/0:1,0:1:3:0,0,0,0
chr01	163799	.	C	T	37.08	PASS	AC=2;AF=	GT:AD:DP:	0/0:1,0:1:3:0,0,0	1 1:0,1:1:3:1 1:	./.:0,0:0:0:0,0,0,0	./.:0,0:0:0:0,0,0,0
chr01	163802	.	G	A	37.08	PASS	AC=2;AF=	GT:AD:DP:	0/0:1,0:1:3:0,0,0	1 1:0,1:1:3:1 1:	./.:0,0:0:0:0,0,0,0	./.:0,0:0:0:0,0,0,0
chr01	174538	.	G	A	40.18	PASS	AC=2;AF=	GT:AD:DP:	1/1:0,2:2:6:49,6	./.:0,0:0:0,0,0,0	./.:0,0:0:0,0,0,0	./.:0,0:0:0,0,0,0
chr01	174549	.	T	A	81.18	PASS	AC=2;AF=	GT:AD:DP:	1 1:0,2:2:6:1 1:	./.:0,0:0:0:0,0,0,0	./.:0,0:0:0:0,0,0,0	./.:0,0:0:0:0,0,0,0
chr01	174554	.	T	C	81.18	PASS	AC=2;AF=	GT:AD:DP:	1 1:0,2:2:6:1 1:	./.:0,0:0:0:0,0,0,0	./.:0,0:0:0:0,0,0,0	./.:0,0:0:0:0,0,0,0
chr01	174564	.	C	T	81.18	PASS	AC=2;AF=	GT:AD:DP:	1 1:0,2:2:6:1 1:	./.:0,0:0:0:0,0,0,0	./.:0,0:0:0:0,0,0,0	./.:0,0:0:0:0,0,0,0
chr01	184582	.	C	T	38.1	PASS	AC=2;AF=	GT:AD:DP:	./.:0,0:0:0:0,0,0,0	./.:0,0:0:0:0,0,0,0	./.:0,0:0:0:0,0,0,0	1 1:0,1:1:3:1 1:184582_C_T
chr01	184583	.	T	C	38.1	PASS	AC=2;AF=	GT:AD:DP:	./.:0,0:0:0:0,0,0,0	./.:0,0:0:0:0,0,0,0	./.:0,0:0:0:0,0,0,0	1 1:0,1:1:3:1 1:184582_C_T
chr01	187743	.	G	A	38.1	PASS	AC=2;AF=	GT:AD:DP:	1/1:0,1:1:3:45,3	./.:0,0:0:0,0,0,0	./.:0,0:0:0,0,0,0	./.:0,0:0:0,0,0,0
chr01	187778	.	T	C	38.1	PASS	AC=2;AF=	GT:AD:DP:	1/1:0,1:1:3:45,3	./.:0,0:0:0,0,0,0	./.:0,0:0:0,0,0,0	./.:0,0:0:0,0,0,0
chr01	187791	.	A	G	38.1	PASS	AC=2;AF=	GT:AD:DP:	1 1:0,1:1:3:1 1:	./.:0,0:0:0:0,0,0,0	./.:0,0:0:0:0,0,0,0	./.:0,0:0:0:0,0,0,0

2. 'IGV'를 통한 변이 검증

- VCF 파일에서 확인된 SNP 변이가 매핑 프로그램의 오류로 발생한 변이인지 실제로 샘플에 존재하는 변이인지 확인이 필요함.
- Bam 파일, 서열파일 등을 입력하면 read mapping 된 정보를 그래픽화 하여 보여주는 'IGV' 프로그램을 통해 2차로 변이를 검증함.
- VCF 파일에서 확인되는 염색체 번호, 변이 위치를 파악하여 결과를 확인함.
- IGV 에 필요한 파일은 아래와 같음.
 - 유전체 서열 (reference.fa / reference.fasta)
 - 유전체 서열의 index 정보 (reference.fa.fai / reference.fasta.fai)
 - 샘플 별 매핑 파일 (sample.sorted.bam)
 - 이때, 매핑 파일은 정렬된 파일이어야 함.
 - 샘플 별 매핑 파일의 index 정보 (sample.sorted.bam.bai)

2. 'IGV'를 통한 변이 검증

□ IGV 프로그램 다운로드 (<https://software.broadinstitute.org/software/igv/home>)

Home | Integrative Genomics V | x +

software.broadinstitute.org/software/igv/

Integrative Genomics Viewer

Home

Downloads

Documents

IGV User Guide

File Formats

Tutorial Videos

Hosted Genomes

Release Notes

Credits

Contact

Search website

search

© 2013-2021 Broad Institute and the Regents of the University of California

Home

Integrative Genomics Viewer

Overview

The Integrative Genomics Viewer (IGV) is a high-performance, easy-to-use, interactive tool for the visual exploration of genomic data. It supports flexible integration of all the common types of genomic data and metadata, investigator-generated or publicly available, loaded from local or cloud sources.

IGV is available in multiple forms, including:

- the original IGV - a Java desktop application,
- IGV-Web - a web application,
- igv.js - a JavaScript component that can be embedded in web pages (for developers)

This site is focused on the IGV desktop application. See <https://igv.org> for links to all forms of IGV.

Download IGV

Download the IGV desktop application and igvtools.

Note that the IGV-Web application at <https://igv.org/app> runs in a web browser and requires no downloads. Click on the Help link in the app for more information.

Citing IGV

To cite your use of IGV in your publication, please reference one or more of:

James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. [Integrative Genomics Viewer. Nature Biotechnology 29, 24-26 \(2011\)](#). (Free PMC article [here](#)).

Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. [Integrative Genomics Viewer \(IGV\): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 14, 178-192 \(2013\)](#).

James T. Robinson, Helga Thorvaldsdóttir, Aaron M. Wenger, Ahmet Zehir, Jill P. Mesirov. [Variant Review with the Integrative Genomics Viewer \(IGV\). Cancer Research 77\(21\) 31-34 \(2017\)](#).

James T. Robinson, Helga Thorvaldsdóttir, Douglass Turner, Jill P. Mesirov. [igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer \(IGV\). Bioinformatics 39\(1\) \(2023\). btac830](#)

Funding

Development of IGV has been supported by funding from the [National Cancer Institute \(NCI\)](#) of the [National Institutes of Health](#), the [Informatics Technology for Cancer Research \(ITCR\)](#), of the NCI, and the [Starr Cancer Consortium](#).

Home | Downloads

Downloads

Did you know that there is also an IGV web application that runs only in a web browser, does not use Java, and requires no downloads? See <https://igv.org/app>. Click on the [Help](#) link in the app for more information about using IGV-Web.

Install IGV 2.16.1

See the [Release Notes](#) for what's new in each IGV release.

Users of the new M1 Mac: Apple's Rosetta software is required to run the IGV MacOS App that includes Java. If you run IGV with your own Java installation, Rosetta may not be required if your version of Java runs natively on M1.

Linux users: The 'IGV for Linux' download includes AdoptOpenJDK (now Eclipse Temurin) version 11 for x64 Linux. See [their list of supported platforms](#). If this does not work on your version of Linux, download the 'Command line IGV for all platforms' and use it with your own Java installation.

About log4j: IGV versions 2.4.1 - 2.11.6 used log4j2 code that is subject to the log4jShell vulnerability. We recommend using version 2.11.9 (or later), which removed all dependencies on log4j.

IGV MacOS App
Java included

IGV MacOS App
Separate Java 11 required

IGV for Windows
Java included

IGV for Windows
Separate Java 11 required

IGV for Linux
Java included

Command line IGV and igvtools for all platforms
Separate Java 11 required

Other IGV Versions

[Development Snapshot Build](#), Latest development snapshot; built at least nightly

[Archived Versions](#), Old releases going back to IGV 2.0

PC 사양에 맞는 버전으로 다운로드
- Java included 버전 권장

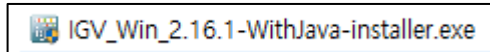
※ 버전에 따라 JAVA 설치가 필요합니다.

JAVA download: <http://www.java.com/ko/download/index.jsp>

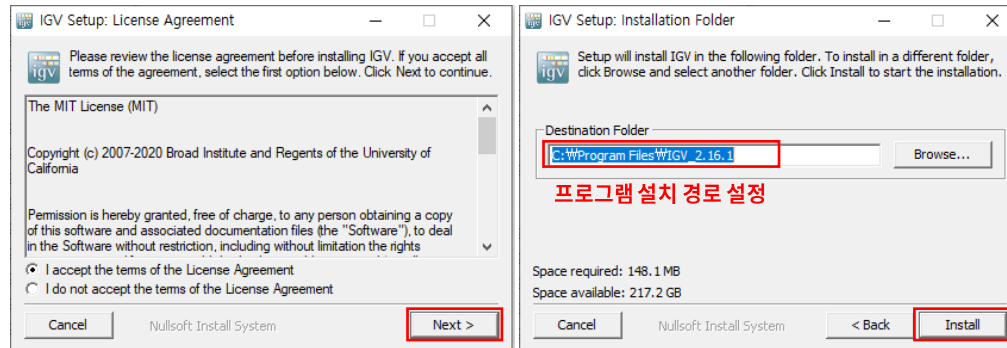
2. 'IGV'를 통한 변이 검증

□ IGV 프로그램 설치

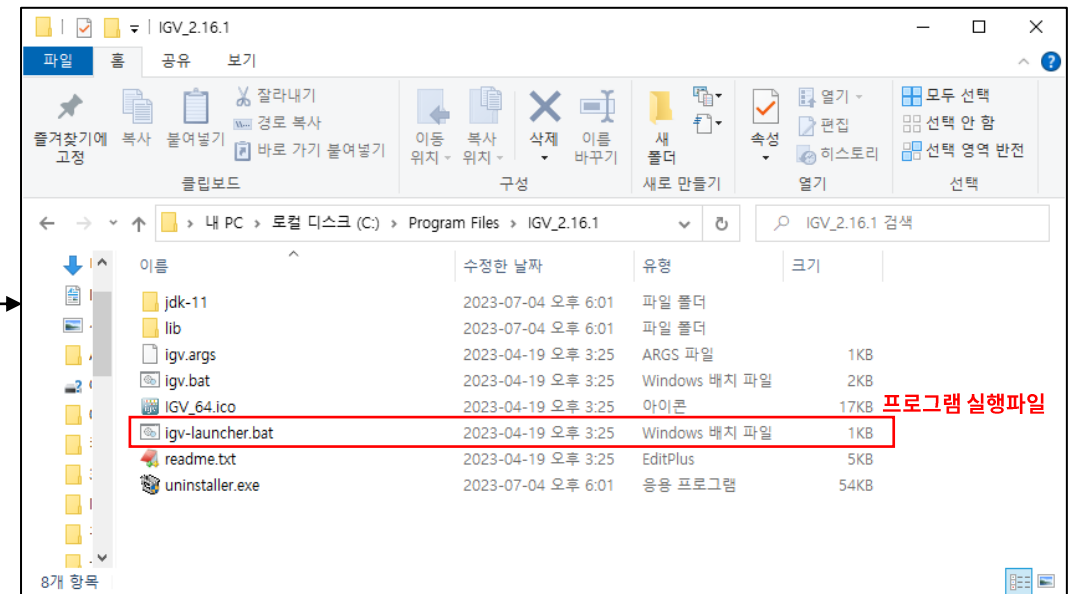
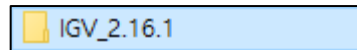
- 다운로드 된 파일



- 다운로드 된 파일 실행



- 설치 완료

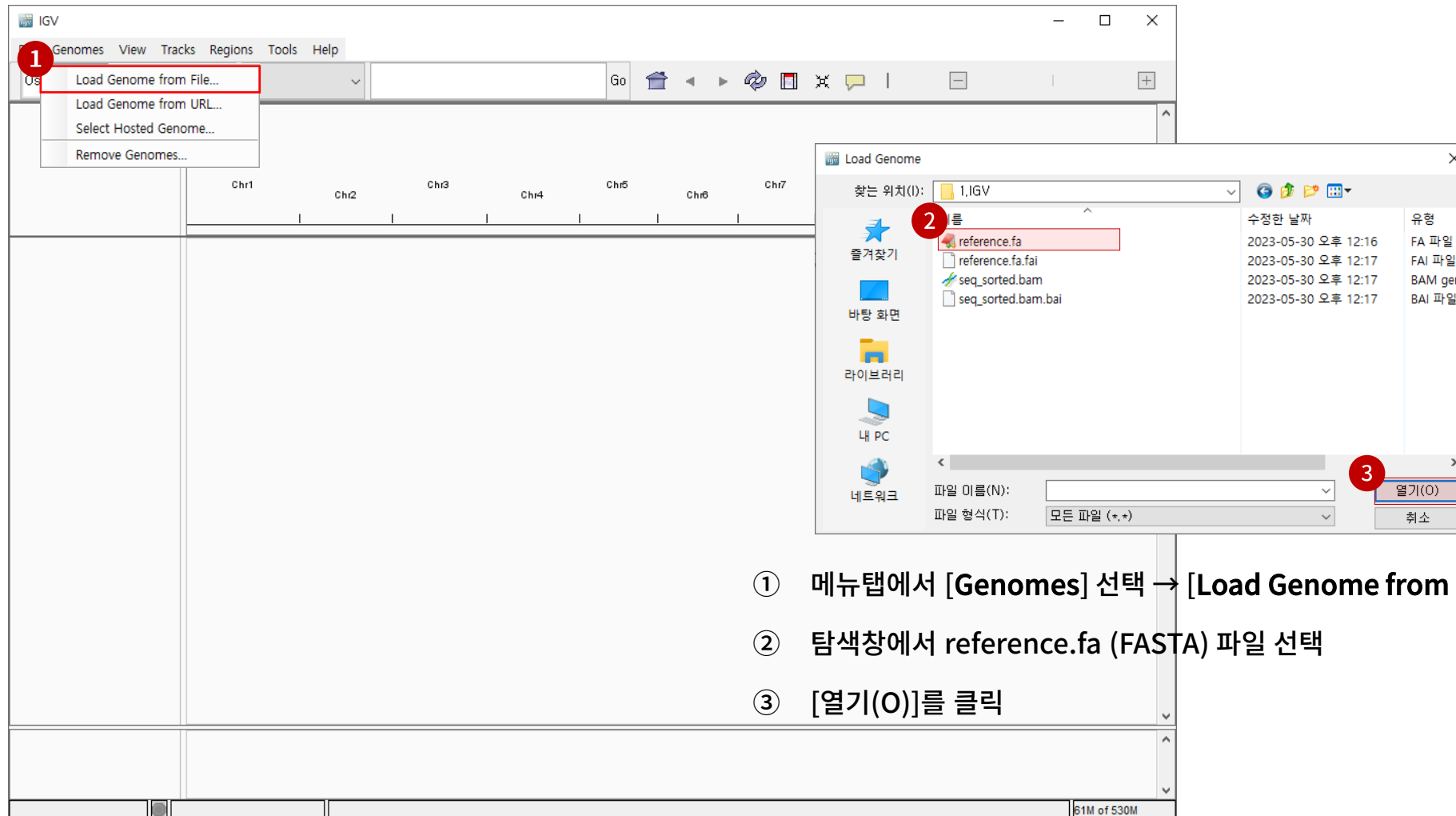


2. 'IGV'를 통한 변이 검증

□ IGV 프로그램 실행 (Genome 파일 등록)

■ 프로그램 실행

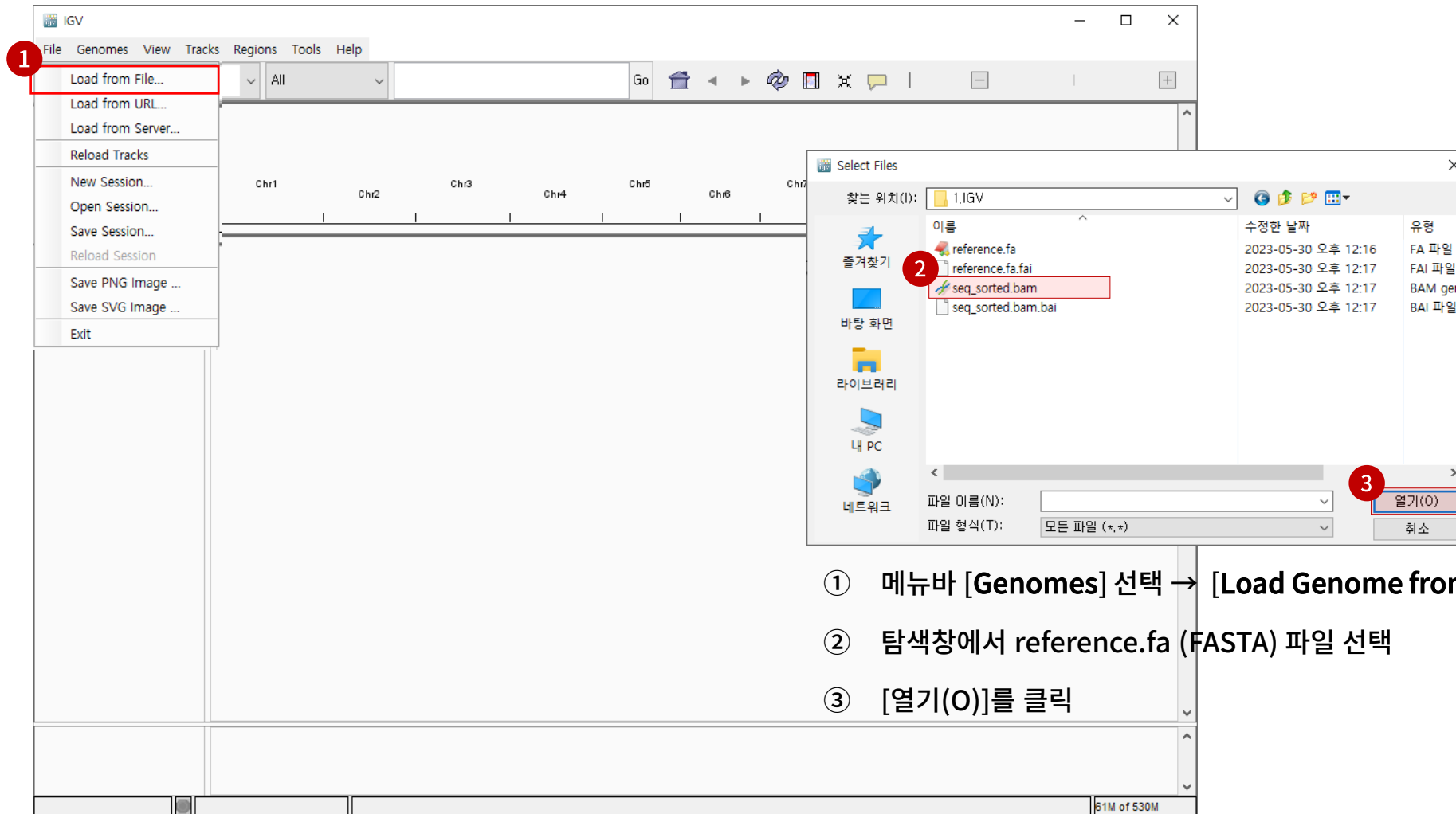
igv-launcher.bat 2023-04-19 오후 3:25 Windows 배치 파일 1KB 프로그램 실행파일



- ① 메뉴탭에서 [Genomes] 선택 → [Load Genome from File...] 선택
- ② 탐색창에서 reference.fa (FASTA) 파일 선택
- ③ [열기(O)]를 클릭

2. 'IGV'를 통한 변이 검증

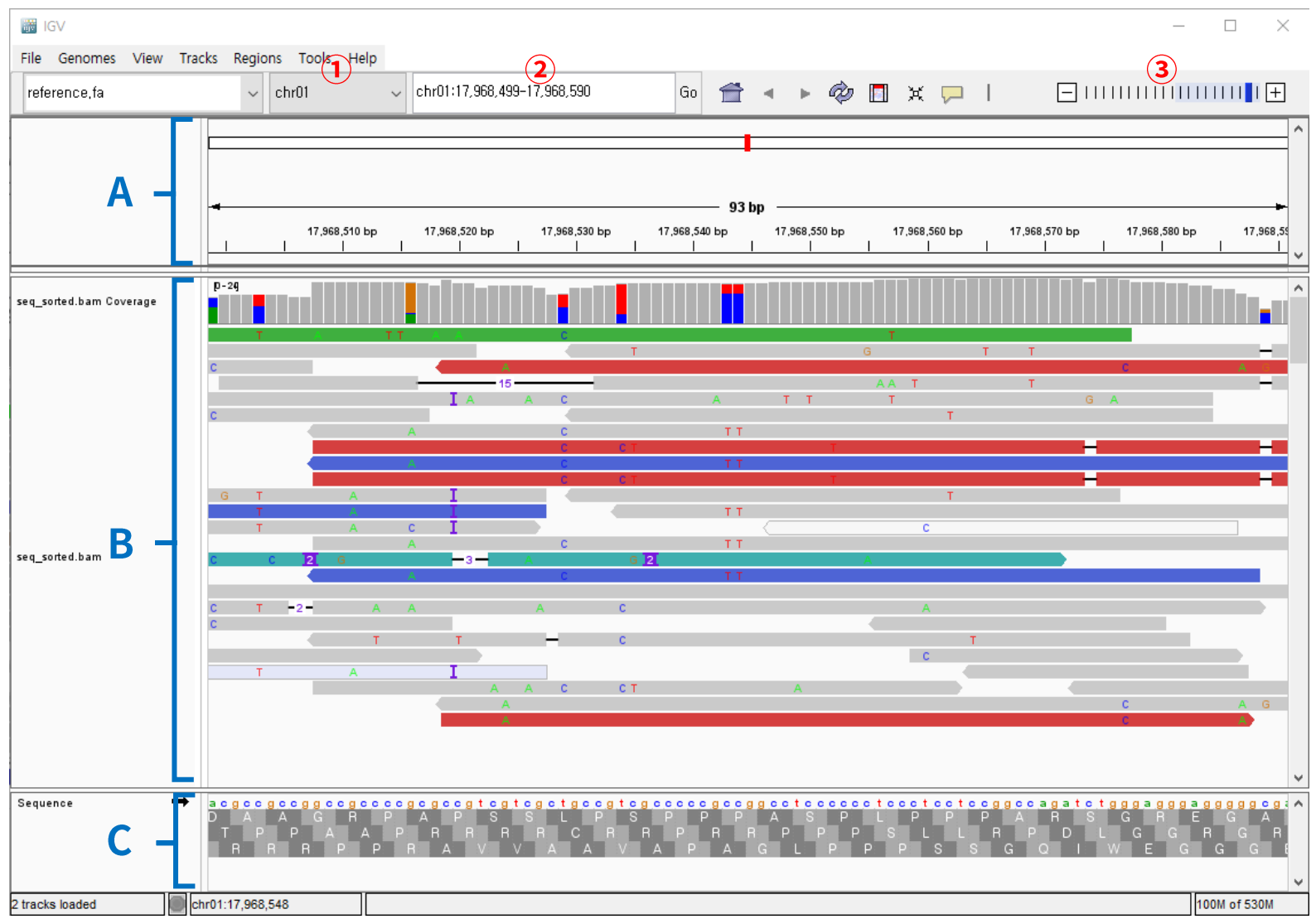
□ IGV 프로그램 실행 (Mapping 파일 등록)



- ① 메뉴바 [Genomes] 선택 → [Load Genome from File...] 선택
- ② 탐색창에서 reference.fa (FASTA) 파일 선택
- ③ [열기(O)]를 클릭

2. 'IGV'를 통한 변이 검증

IGV 메뉴 설명

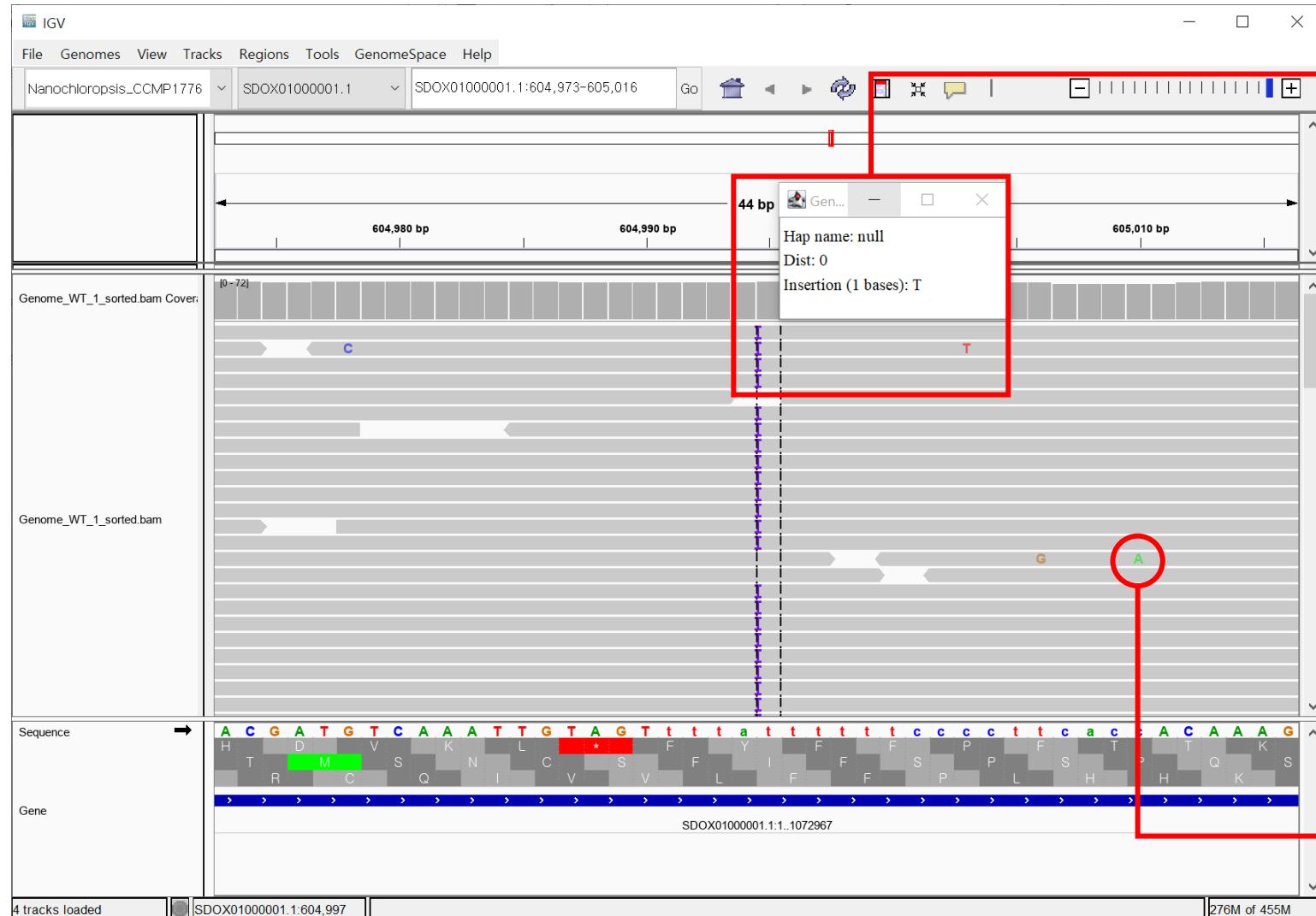


- ① 염색체 선택
- ② 위치 정보 (염색체:시작-끝) 정보 입력 시 해당 위치로 이동 가능
- ③ 유전체 영역 확대(+) 및 축소(-)

- A. 현재 보여지는 유전체 영역의 위치 및 크기
- B. 샘플 별 read alignment 정보
- C. Reference의 염기서열, 아미노산 서열, 유전자 정보 (gff 파일 입력시 확인 가능)

2. 'IGV'를 통한 변이 검증

□ Alignment & Variants 확인



❖ Mapped reads 에서 insertion이 발생할 경우 |으로 표시되고, 클릭하면 팝업창으로 삽입되는 염기서열을 확인할 수 있음.

Mapped coverage

Mapped reads

Reference bases

※ Mapped reads 에서 reference base와 다른 경우, mismatch 되는 염기서열을 표시함.

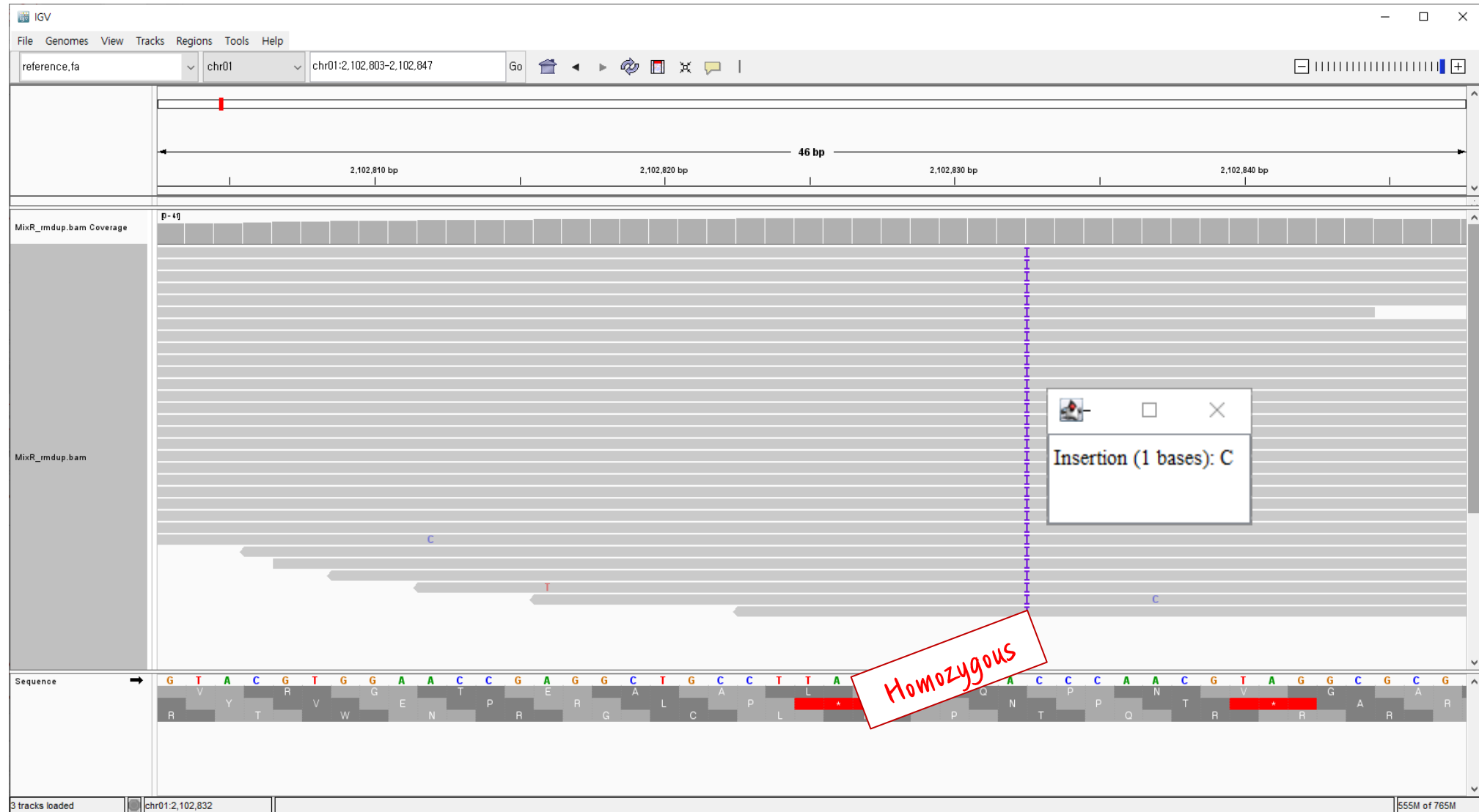
2. 'IGV'를 통한 변이 검증

□ Variants 확인 – Homo/heterozygous SNP



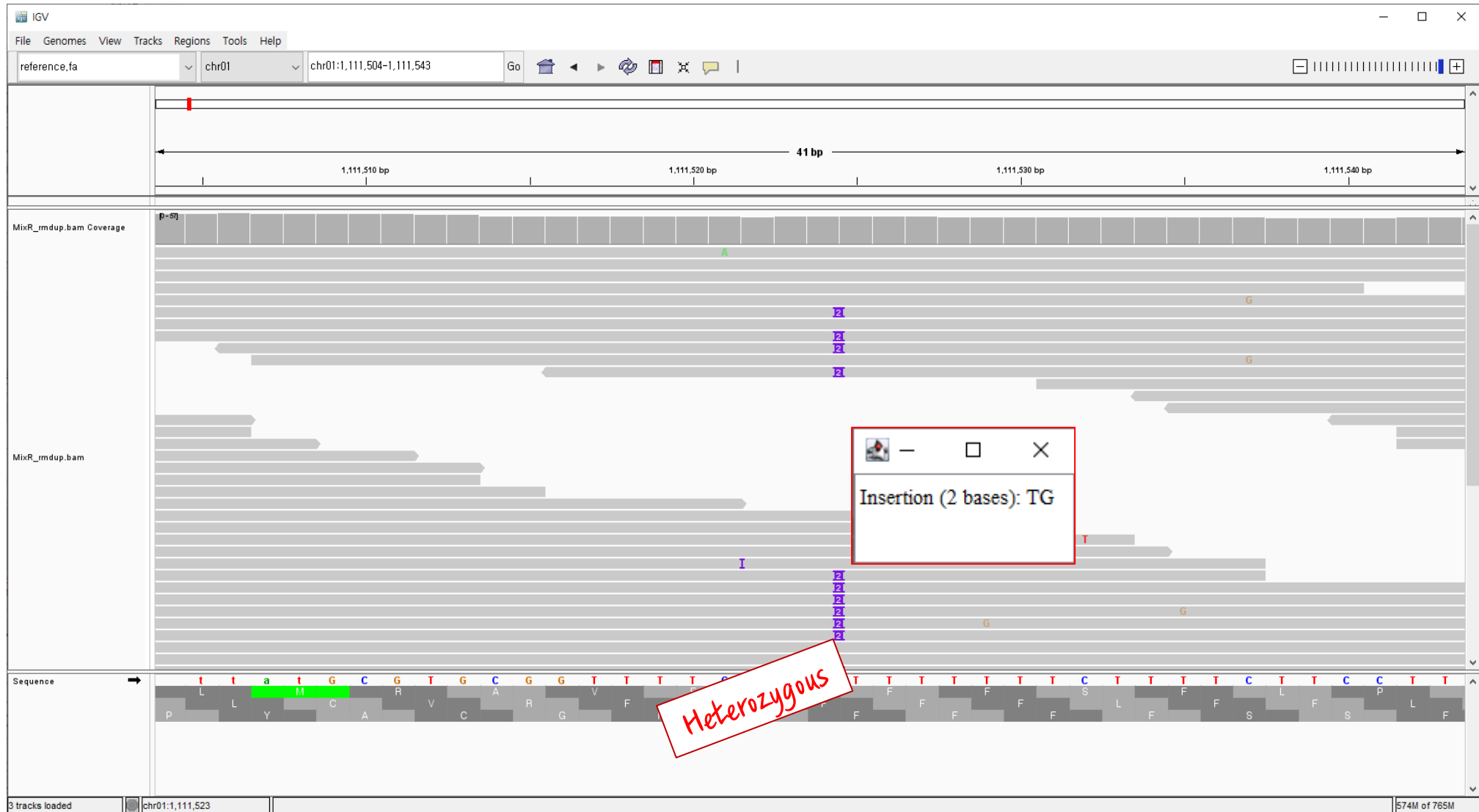
2. 'IGV'를 통한 변이 검증

□ Variants 확인 – Homozygous Insertion



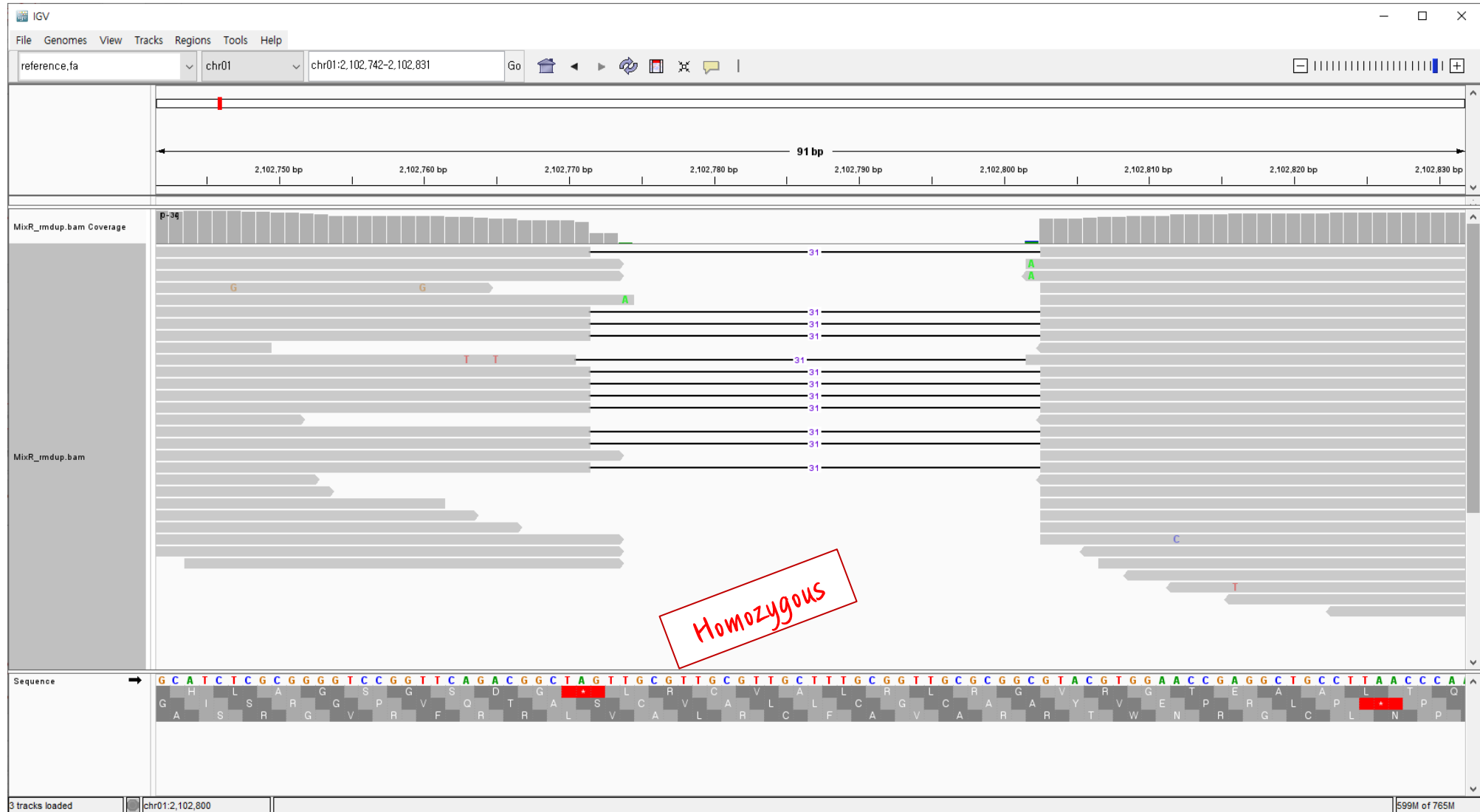
2. 'IGV'를 통한 변이 검증

□ Variants 확인 - Heterozygous Insertion



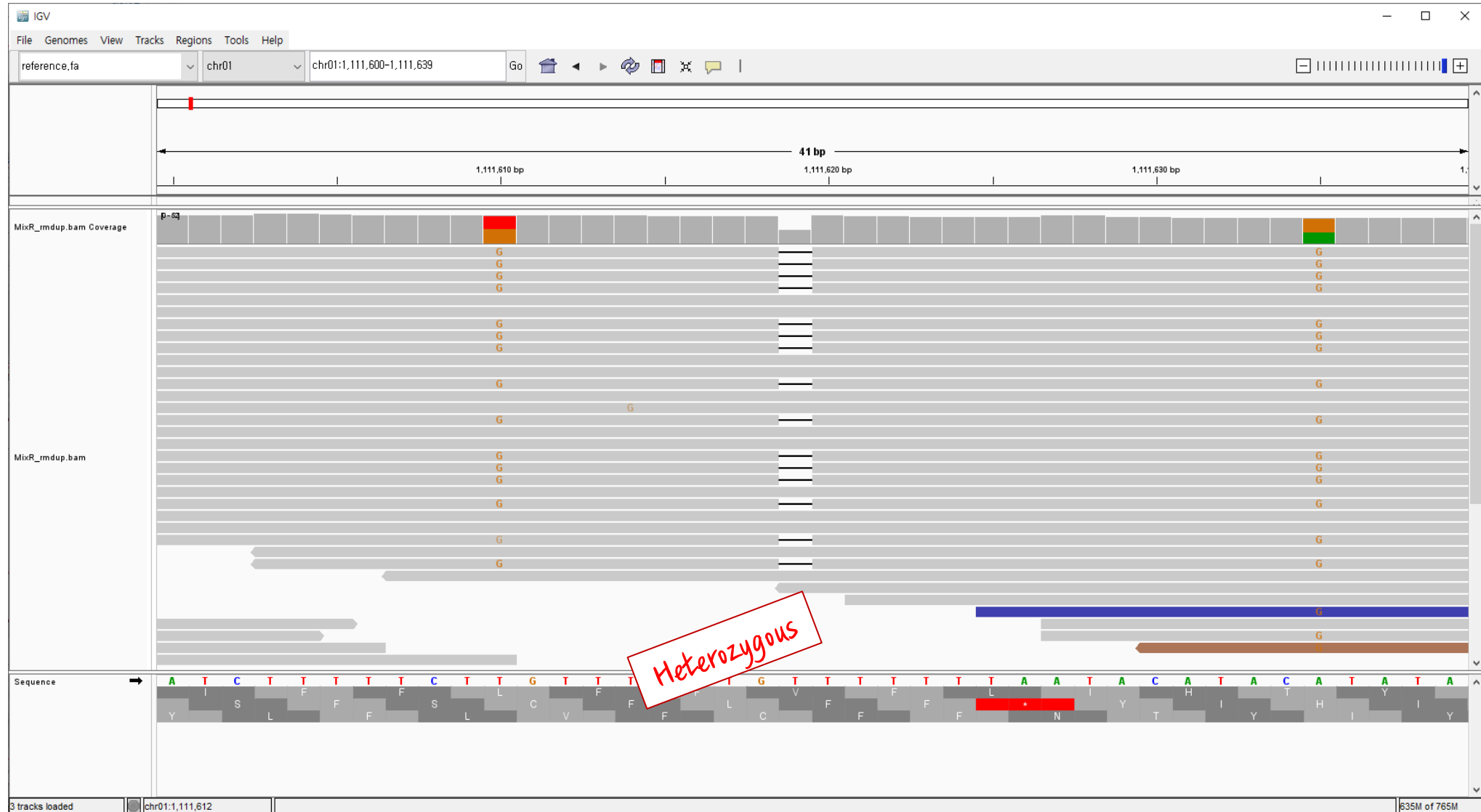
2. 'IGV'를 통한 변이 검증

□ Variants 확인 – Homozygous Deletion



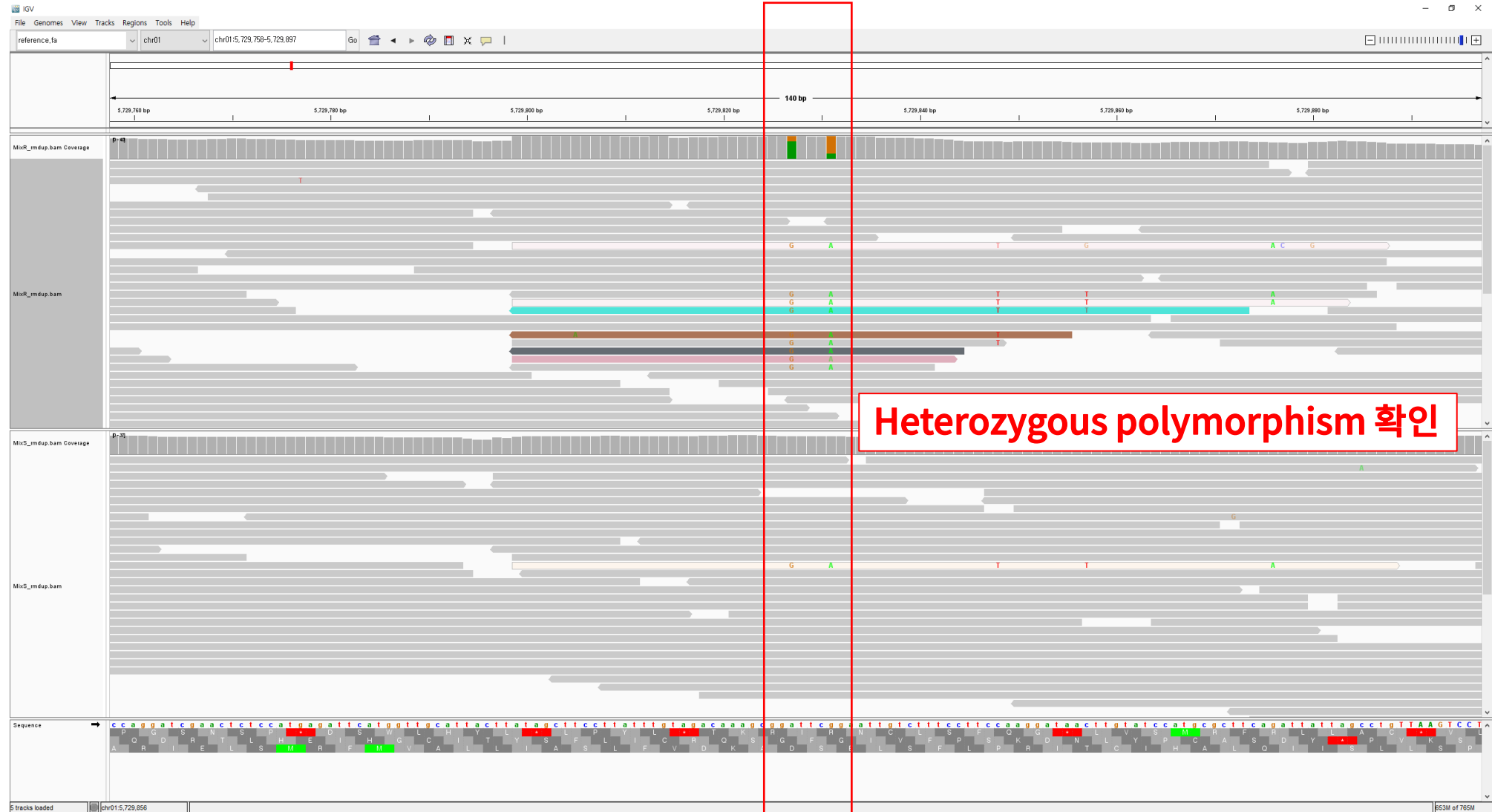
2. 'IGV'를 통한 변이 검증

□ Variants 확인 - Heterozygous Deletion



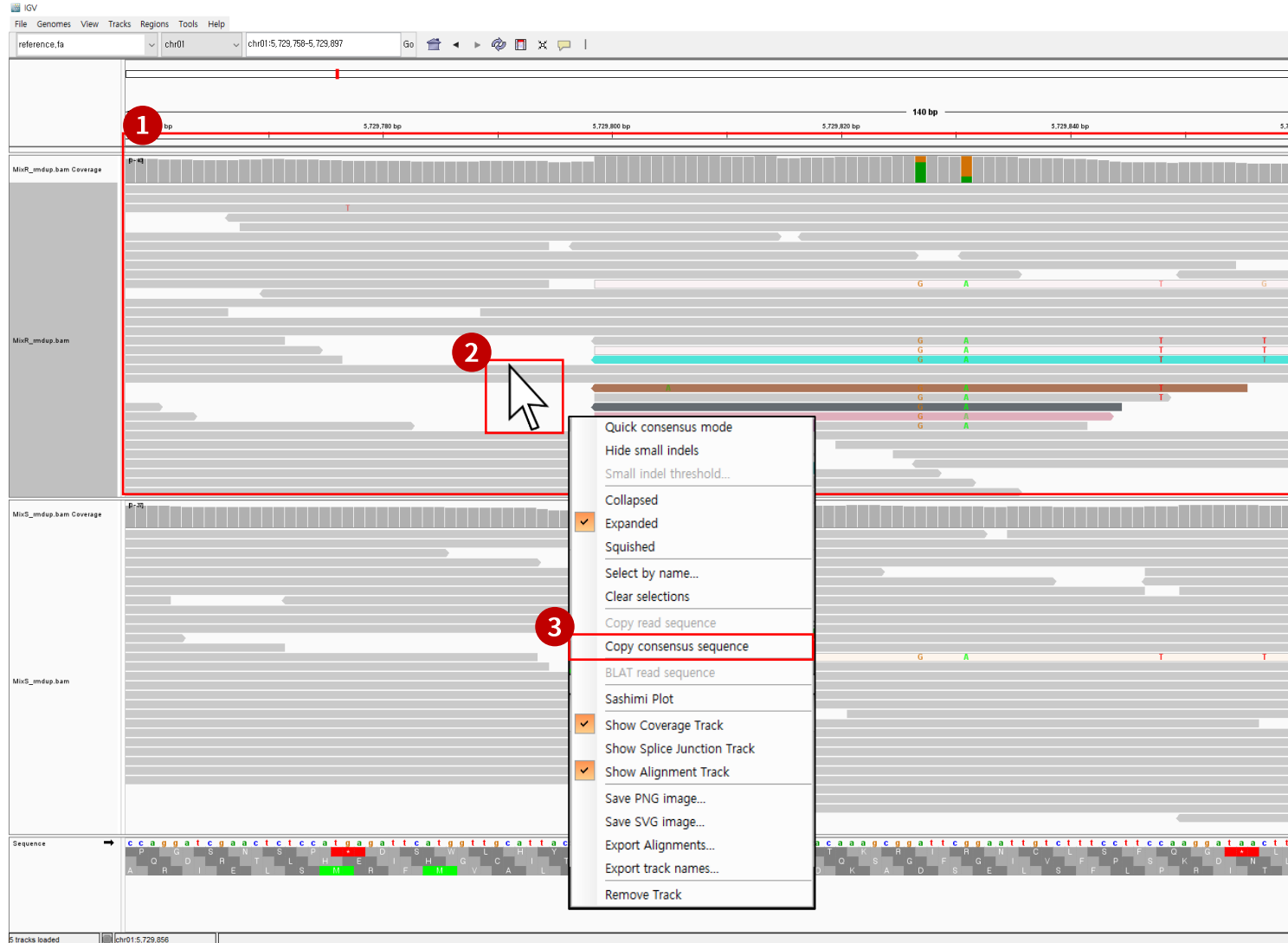
3. Polymorphism 확인

□ 'IGV' 프로그램에서 비교 샘플 간의 polymorphism을 확인



4. 변이 주변 서열 확보

□ 'IGV' 프로그램에서 대상 변이 (SNP)의 주변 서열 확보 (약 500bp 영역 선발)



```

4
chr01:5729553-5730110
GATATGATTGACAGGAAGTGCAATGACATGTCCTTACACCAACAG
GATGTGATAAAGATAATGAAGCTTGCAATGTGGTGCTTGCAGAGT
GATTGCAACAGAAGGCCTTCCATGTCATTGGTGATGAAGTCCTT
GAAGGAGAAAAGTGATGTTGAAGCCAACCTTGAACATAACTTCTTT
GATTGCCGCCAGCGTTCATCCTGAGCCAGGATCGAACTCCTCATG
AGATTCATGGTTGCATTACTTATAGCTTCCTTATTGTAGACAAA
GCGGATTCGGAATTGTCTTTCCTTCCAAGGATAACTTGTATCCAT
GCGCTTCAGATTATTAGCCTGTTAAGTCCGCGAATTATGTTCC
GTTGACCCATTCAAGATGTCACCTGCATGGCTCCATGCCTCCACCT
GAAGCAAATGCAGATGTAGAAAACCATTTTGTGAACCTGTATTAC
TGCAATCATATCATTGATATCACACATCGATCAGTGCATACCCAA
GTTAATCTGTGATTGACTAGAAATGGTAGTACTGTTTCAGTTCA
GTTATTTTGTGCTTTTTG
    
```

- ① 확보하고자 하는 변이 주변 서열로 확대
- ② 샘플 별 read mapping 결과 창에서 마우스 오른쪽 클릭
- ③ Copy read sequence 선택하여 샘플의 consensus 서열 추출
- ④ 메모장에 ctrl+v 로 붙여넣기

IGV 사용법 참고자료

□ IGV 프로그램에 대한 자세한 사용법은 IGV User Guide 를 통해 확인할 수 있음.

(<http://software.broadinstitute.org/software/igv/UserGuide>)

Home > IGV User Guide

IGV User Guide

This guide describes the Integrative Genomics Viewer (IGV).

- To start IGV, go to the IGV downloads page: <http://www.broadinstitute.org/igv/download>

[Look at a printer-friendly HTML version of the whole User Guide.](#)

- [User Interface](#)
- [Navigating the View](#)
- [Loading a Genome](#)
- [External Control of IGV](#)
- [Viewing the Reference Genome](#)
- [Loading Data and Attributes](#)
- [Viewing Data](#)
- [Viewing Alignments](#)
- [Viewing Variants](#)
- [Multi-Locus View](#)
- [Regions of Interest](#)
- [Sample Attributes](#)
- [Sorting, Grouping, and Filtering](#)
- [Saving and Restoring Sessions](#)
- [Server Configuration](#)
- [Motif Finder](#)
- [igytools](#)
- [BLAT search](#)

[User Interface >](#)

Part II. Primer 디자인 및 결과 확인

1. 'Primer3'를 통한 PCR primer design

□ 'Primer3' 프로그램 수행하기 앞서 input 파일 작성을 진행

- Primer3 input 파일 형식

```

SEQUENCE_ID=<SNP_ID>
SEQUENCE_TEMPLATE=<Sequence>
SEQUENCE_TARGET=<Target position, target size>
PRIMER_NUM_RETURN=<Primer 개수>
PRIMER_EXPLAIN_FLAG=1
PRIMER_TASK=pick_pcr_primers
PRIMER_PICK_LEFT_PRIMER=1
PRIMER_PICK_INTERNAL_OLIGO=0
PRIMER_PICK_RIGHT_PRIMER=1
PRIMER_PRODUCT_SIZE_RANGE=<예상 Product size>
PRIMER_MIN_SIZE=18
PRIMER_OPT_SIZE=20
PRIMER_MAX_SIZE=24
PRIMER_MIN_TM=55
PRIMER_OPT_TM=60
PRIMER_MAX_TM=65
PRIMER_MAX_NS_ACCEPTED=0
PRIMER_THERMODYNAMIC_PARAMETERS_PATH=<Primer3 program path>
=
    
```

1. 'Primer3'를 통한 PCR primer design

□ 'Primer3' 프로그램 수행하기 앞서 input 파일 작성을 진행

- 파일명: PCR_primer3_input.txt

```

SEQUENCE_ID=chr01:5729553-5730110
SEQUENCE_TEMPLATE=GATATGATTGACAGGAAGTGCAATGACATGTCCTTACACCAACAGGATGTGATAAAGATAATGAAGCTTGCAATGTGGTGCTTGCA
GAGTGATTGCAACAGAAGGCCTTCCATGTCATTGGTGATGAAGGTCCTTGAAGGAGAAAAGTGATGTTGAAGCCAACCTTGAACATAAATTCTTTGATTGCCGCC
AGCGTTCATCCTGAGCCAGGATCGAACTCTCCATGAGATTCATGGTTGCATTACTTATAGCTTCCTTATTTGTAGACAAAGCGGATTCGGAATTGTCTTTCCTT
CCAAGGATAACTTGTATCCATGCGCTTCAGATTATTAGCCTGTAAAGTCCTGCGAATTATGTTCCCGTTGACCCATTCAAGATGTCACACTGCATGGCTCCATGCC
TCCACCTGAAGCAAATGCAGATGTAGAAAACCATTTTTGTGAACTGTATTACTGCAATCATATCATTGATATCACACATCGATCAGTGCATACCCAAGTTAATC
TGTGTATTGACTAGAAATGGTAGTACTGTTTCAGTTCAGTTATTTTTGTGTCTTTTG
SEQUENCE_TARGET=274,10
PRIMER_EXPLAIN_FLAG=1
PRIMER_TASK=pick_pcr_primers
PRIMER_PICK_LEFT_PRIMER=1
PRIMER_PICK_INTERNAL_OLIGO=0
PRIMER_PICK_RIGHT_PRIMER=1
PRIMER_PRODUCT_SIZE_RANGE=100-200
PRIMER_MIN_SIZE=18
PRIMER_OPT_SIZE=20
PRIMER_MAX_SIZE=24
PRIMER_MIN_TM=55
PRIMER_OPT_TM=60
PRIMER_MAX_TM=65
PRIMER_MAX_NS_ACCEPTED=0
PRIMER_THERMODYNAMIC_PARAMETERS_PATH=/agribio/HOME/edu_02/prepare/dataset/primer3_config/
=
    
```


1. 'Primer3'를 통한 PCR primer design

□ 'Primer3' 프로그램 수행하기 앞서 input 파일 작성을 진행

- Primer input 파일 생성

```
vi <파일명> # 파일명: PCR_primer3_input.txt
```

- 파일 편집하기

```
파일 편집하기 : i 키 입력
파일 저장 후 닫기 : esc 키 입력 → shift + : → wq!
```



```
~/
~/
~/ vi PCR_primer3_input.txt
```



1. 'Primer3'를 통한 PCR primer design

□ 'Primer3' 프로그램 수행

- Primer3 실행 명령어

```
primer3_core -format_output -strict_tags -output=<output파일> <input파일>
```

실행명령어 1: primer3_core -format_output -strict_tags -output=PCR_primer3_result.txt
PCR_primer3_input.txt

- Primer3 결과 파일 확인하기

```
less PCR_primer3_result.txt
```


2. ‘Bowtie’를 PCR primer 검증

- Primer3를 통해 작성된 PCR primer 후보들은 표준유전체 서열 내 중복으로 존재하는지, *in-silico* 상에서 정상적으로 분석이 되는지 검증이 필요함.
- PCR primer 서열의 검증은 표준유전체 서열에 작성된 primer 정보로 read alignment 를 수행하여 검증을 진행함.
- *in-silico* PCR 은 read align 이 비교적 빠른 시간내 수행되는 ‘bowtie’ tool 을 이용함.
- PCR primer 검증은 아래와 같이 총 3단계를 걸쳐 수행함.
 - bowtie input 파일 작성
 - bowtie align 수행
 - in-silico PCR 결과 확인 및 unique PCR primer 선별

2. 'Bowtie'를 PCR primer 검증

□ 'Bowtie' input 파일 작성

- Input 파일 작성하기

```
vi <파일명>
```

- 파일 편집하기

파일 편집하기 : i 키 입력

파일 저장 후 닫기 : esc 키 입력 → shift + : → wq!



파일명: PCR_primer_F.fa

```
>primer1_F
AGCCAGGATCGAACTCTCCA
>primer2_F
GAGCCAGGATCGAACTCTCC
```

파일명: PCR_primer_R.fa

```
>primer1_R
GGAGCCATGCAGTGACATCT
>primer2_R
ATGGAGCCATGCAGTGACAT
```

2. 'Bowtie'를 PCR primer 검증

□ 'Bowtie' 명령어 작성 및 수행

- Bowtie index 작성 명령어

```
bowtie-build <reference 서열> <index 파일명>
```

실행명령어 2: `bowtie-build reference.fa reference.fa`

- Bowtie 실행 명령어

```
bowtie -a -p <cpu 개수> -f -X 500 --fr <reference 서열> -1 <Primer forward 서열> -2 <Primer reverse 서열> <alignment 결과 파일>
```

[옵션 설명]

-a	모든 read의 alignment 된 정보 보기
-p	thread (CPU) 개수
-f	입력하는 파일 타입 (fasta -> f, fastq -> q)
-X	paired end 서열 내 최대 insert size
--fr	-1, -2에 입력하는 파일은 forward/reverse로 align
-1	forward 파일명
-2	reverse 파일명

실행명령어 3: `bowtie -a -p 5 -f -X 500 --fr reference.fa -1 PCR_primer_F.fa -2 PCR_primer_R.fa PCR_primer_result.aln`

3. Unique PCR primer 선발

□ 'Bowtie' 수행 결과에서 unique 한 PCR primer 선발을 진행

- 실행 후 출력창

```
# reads processed: 2
# reads with at least one reported alignment: 2 (100.00%)
# reads that failed to align: 0 (0.00%)
Reported 2 paired-end alignments
```

- 이때, 'unique 하다'는 유전체 서열에 PCR primer 서열이 1번만 존재하는 것을 의미함.
- *in-silico* PCR 결과 열어보기

```
less <파일명> # 파일명: PCR_primer_result.aln
```

primer1_F/1	+	chr01	5729755	AGCCAGGATCGAACTCTCCA	IIIIIIIIIIIIIIIIIIIIII	0
primer1_R/2	-	chr01	5729925	AGATGTCACTGCATGGCTCC	IIIIIIIIIIIIIIIIIIIIII	0
primer2_F/1	+	chr01	5729754	GAGCCAGGATCGAACTCTCC	IIIIIIIIIIIIIIIIIIIIII	0
primer2_R/2	-	chr01	5729927	ATGTCACTGCATGGCTCCAT	IIIIIIIIIIIIIIIIIIIIII	0



감사합니다.



대전시 유성구 테크노2로 187, B동 412호



bi@bioto.co.kr



042-710-0077



070-7585-5344