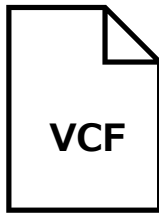
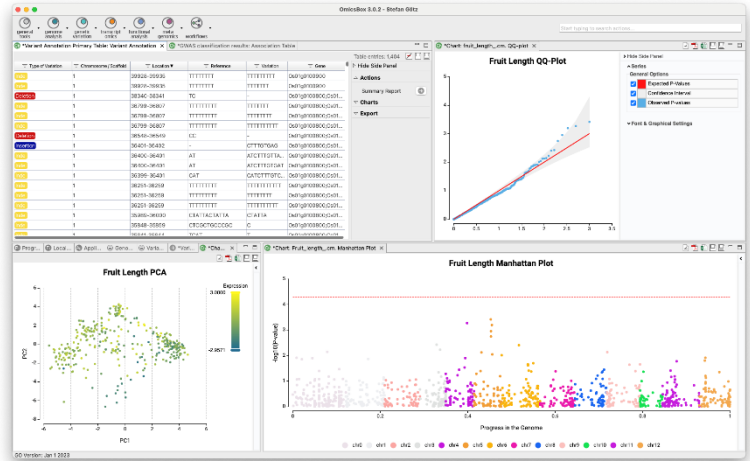


# Genetic Variation

## Fast Variant Calling

BAMファイルを使用した変異検出、フィルタリング、およびアノテーションを実行できるだけでなく、ゲノムワイド関連解析 (GWAS)を介して遺伝的変異を特定の形質または疾患に関連付けることができます (ショートリードデータのみ対応)。特定のメソッド (例: GBS、WGS) や要件 (倍数性レベル、ゲノムカバレッジ、反復領域の有無など) に合わせて解析を調整するための多くのオプションが利用できます。

\*リファレンスゲノムへのマッピングは、別途 Genome Analysis Module が必要です。



## Variant Calling

マッピングデータからバリエーション (SNP、insertion、または deletionsなど) を識別するプロセスです。

Chromosome	Position	Reference	Variant	Quality	Filter
1	221932	T	C	10	PASS
1	221932	T	C	10	PASS
1	221932	T	C	10	PASS
1	221932	T	C	10	PASS
1	221932	T	C	10	PASS
1	221932	T	C	10	PASS
1	221932	T	C	10	PASS
1	221932	T	C	10	PASS
1	221932	T	C	10	PASS
1	221932	T	C	10	PASS

## Variant アノテーション

Ensembl の Variant Effect Predictorを使用して、VCF ファイルに含まれるすべてのバリエーションのコーディングと遺伝的影響に関する情報を収集します。必要なのは Variant Calling/Filtering ステップで使用したリファレンスゲノムとアノテーションファイル(.gtf)だけです。

## Variant Filtering Report

**Input Data**  
freebayes.vcf.gz

**Results**

Number of variants before filtering	3903	Number of variants after filtering	2001	Percentage filtered	48.7291%
-------------------------------------	------	------------------------------------	------	---------------------	----------

**Parameters**

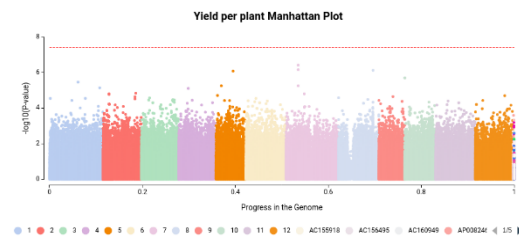
Parameter	Value
Proportion Quality / Coverage	2.0
Raw Read Depth	5
Check Reads in Both Strands	true
Check if Reads are Balanced	true
Average Mapping Quality	50.0
Phred Quality	1.0
Remove Multiple Alleles	true

**References**

- Danecek P et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience, 10(2).
- Genious - Bioinformatics made easy. Bioinformatics (version 3.0.2). March 3, 2019. www.bioinformatics.com

## Variant フィルタリング

VCFファイルのさまざまなフィールドに関する独自の基準に従って、品質の低いバリエーションを削除します。



## GWAS

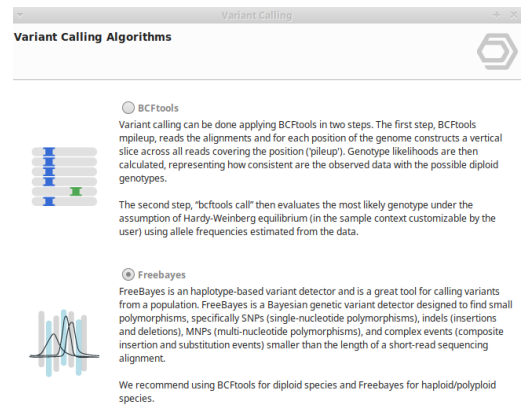
GAIT3 Rパッケージを使用した、ゲノムワイド関連研究 (GWAS) により、多くのゲノムに存在する数千から数百万の遺伝子変異を検査し、特定の形質に関連する遺伝的変異を特定できます。



# Genetic Variationモジュールでなにができる？

## 【使用したデータセット】

このモジュールでは、一般的なツールであるBCFtoolsとFreeBayesに基づくバリエーションコールとフィルタリングのための2つの異なる分析ストラテジーを提供します。特定の方法論（例：GBS vs WGS）や要件（倍数性のレベル、ゲノムカバレッジ、反復領域の有無など）に合わせて解析を調整するための多くのオプションが利用可能です。得られたVCFファイルは、EnsemblのVariant Effect Predictorを使用してアノテーションすることができます。このツールの組み合わせは、最近のいくつかのレビュー研究において、代替パイプラインを凌駕することが示されています（e.g. Nature Sci. Rep. 12, 11331 (2022)）。



**Variant Calling Algorithms**

- BCFtools**  
Variant calling can be done applying BCFtools in two steps. The first step, BCFtools mpileup, reads the alignments and for each position of the genome constructs a vertical slice across all reads covering the position (pileup). Genotype likelihoods are then calculated, representing how consistent are the observed data with the possible diploid genotypes.
- Freebayes**  
FreeBayes is an haplotype-based variant detector and is a great tool for calling variants from a population. FreeBayes is a Bayesian genetic variant detector designed to find small polymorphisms, specifically SNPs (single-nucleotide polymorphisms), indels (insertions and deletions), MNPs (multi-nucleotide polymorphisms), and complex events (composite insertion and substitution events) smaller than the length of a short-read sequencing alignment.

We recommend using BCFtools for diploid species and Freebayes for haploid/polyploid species.

## 【最終的に得られる結果】

Variant Calling→Variant フィルタリングの解析の後、実行したバリエーションコール結果では、見つかった各バリエーションの情報を含むテーブルといくつかの集団遺伝情報に関する情報を含むレポートが得られます。以下はバリエーションの情報を含むテーブルと結果で、変異のタイプのタグ、変異として VCF に表示されるヌクレオチドまたは配列、Nucleotide diversity(集団内の遺伝的多様性のレベル、表中のPi)、HWE p-valueが含まれます。さらに各バリエーションの結果の詳細を閲覧することも可能です(Annotation Detailsのテーブル)。この詳細にはその変異によって影響を受ける遺伝子機能の名前（または ID）、cDNA・CDS・Proteinの位置、アミノ酸置換情報が含まれます。また、得られた結果からTs/Tv 比の情報のレポート表示や変異タイプ分布の円グラフ、品質管理チャートなどの作図などが可能です。

Type of Variation	Chromosome / Scaffold	Location	Reference	Variation	Gene	Pi	HWE p-value
Insertion	4	2709735-2709736	-	GATCATCC	-	0.0	1.0
Indel	5	2594922-2594923	CT	CTACT	Os05g0144800;Os05g014490...	0.284266	0.3317609
Indel	8	2397155-2397158	CCAT	CCCATCCAT	Os08g0143500	0.0204082	1.0
Indel	8	2397155-2397158	CCAT	CCATCCAT	Os08g0143500	0.0204082	1.0
Indel	2	1965696-1965698	CAA	ACAA	Os02g0137000;Os02g0137100	0.0	1.0
Deletion	2	1965696-1965698	CAA	-	Os02g0137000;Os02g0137100	0.0	1.0
Indel	2	1965696-1965698	CAA	CCCACAA	Os02g0137000;Os02g0137100	0.0	1.0
Indel	6	1646252-1646257	CTTTCT	CCTTTCT	Os06g0130800;Os06g013090...	0.0	1.0
Indel	6	1646252-1646257	CTTTCT	CT	Os06g0130800;Os06g013090...	0.0	1.0
Indel	6	1368940-1368943	TGCT	T	Os06g0125300;Os06g012540...	0.364731	0.1389143
Indel	5	1158632-1158636	AGCTA	A	Os05g0121600	0.447689	0.06195259
Indel	5	1158632-1158636	AGCTA	CTAGCTA	Os05g0121600	0.447689	0.06195259
Indel	9	1072045-1072047	CAC	CACCCAC	-	0.0493671	1.0
Indel	12	976125-976127	AAA	AAAA	Os12g0120400;Os12g0120500	0.162319	1.0
Indel	1	481796-481805	GAGAGGAGAG	G	Os01g0108600;Os01g010870...	0.0163934	1.0
Deletion	11	283552-283569	GCTATAGGAGAAAGCCAAA	-	Os11g0106083;Os11g010610...	0.483689	0.823286
Insertion	7	178003-178004	-	A	Os07g0103000	0.0416667	1.0
SNP	1	43269831	T	A	-	0.197712	1.0
SNP	1	43269411	T	G	-	0.0	1.0

バリエーションの情報を含むテーブル

## Annotation Details

Gene: Os06g0130800

Feature ID	Feature Type	Consequence	cDNA Position	CDS Position	Protein Position	Amino Acids	Codons	Impact	Distance	Strand
Os06t0130800-01	Transcript	upstream_gene_variant	-	-	-	-	-	MODIFIER	2453	-1

Gene: Os06g0130900

Feature ID	Feature Type	Consequence	cDNA Position	CDS Position	Protein Position	Amino Acids	Codons	Impact	Distance	Strand
Os06t0130900-00	Transcript	frameshift_variant	66-71	65-70	22-24	EKA/EKGX	gAGAAAGcc/gAGAAAGGcc	HIGH	-	-1

Gene: Os06g0131001

Feature ID	Feature Type	Consequence	cDNA Position	CDS Position	Protein Position	Amino Acids	Codons	Impact	Distance	Strand
Os06t0131001-00	Transcript	frameshift_variant	567-572	567-572	189-191	GFL/GLSX	ggCTTTCT/ggCCTTTCTc	HIGH	-	1

Gene: Os06g0131100

Feature ID	Feature Type	Consequence	cDNA Position	CDS Position	Protein Position	Amino Acids	Codons	Impact	Distance	Strand
Os06t0131100-01	Transcript	upstream_gene_variant	-	-	-	-	-	MODIFIER	2338	1
Os06t0131100-02	Transcript	upstream_gene_variant	-	-	-	-	-	MODIFIER	2360	1

各バリエーションの結果の詳細（Annotation Details）

Genetic Variationモジュールのツールの組み合わせは、最近のレビュー研究で他の方法よりも優れた方法で、遺伝子変異解析を実行できます！