

# OmicsBox with Genome Analysis

Genome의 특성을 효율적이고 사용자 친화적인 방식으로 특성화할 수 있는 Tool



## GENOME ANALYSIS

- Quality control and assessment
- De novo assembly
- Repeat masking
- Gene finding

## Quality control and assessment

샘플의 품질 관리를 수행하기 위해, FastQC와 Trimmomatic을 사용하여 Reads를 필터링하고 low quality bases를 제거할 수 있습니다.

## De novo assembly

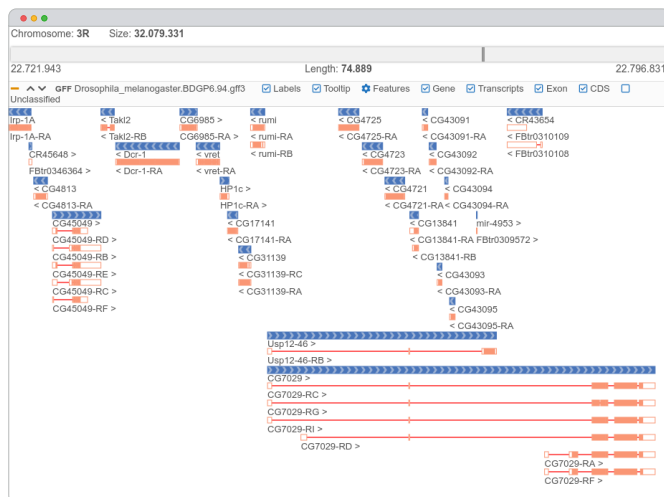
ABYSS를 기반으로 하는 assembly 기능을 통해 reference genome 또는 특정 하드웨어 요구 사항 없이 전체 genome sequence를 재구성할 수 있습니다.

## Repeat masking

다운스트림 유전자 예측을 개선하기 위해 RepeatMasker를 사용하여 진핵생물 genome의 반복적이고 복잡성이 낮은 assemble된 DNA sequence를 마스킹 할 수 있습니다.

## Genome browser

트랙 형태로 annotation을 시각화하여 genome sequences (.fasta), alignments (.bam), intron-exon structure (.gff) 및 variant data (.vcf) 와 결합할 수 있습니다.



## Gene finding

Genome 구조를 특성화하기 위해 원핵생물 (Glimmer) 및 진핵생물 (Augustus) 유전자 예측을 수행할 수 있으며, 진핵생물 유전자 예측은 RNA-seq 인트론 힌트를 지원합니다.

## Exploratory analysis

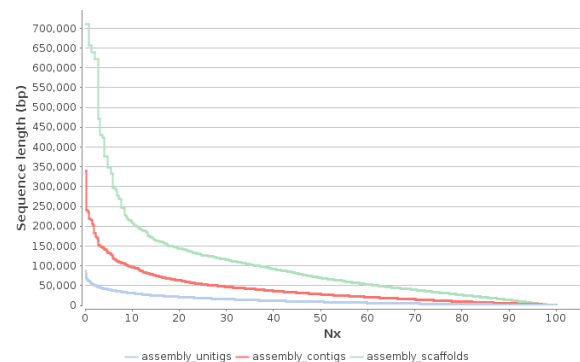
풍부한 사용자 인터페이스를 통해 대규모 genome annotation을 쉽게 처리할 수 있고, General Feature Format 스타일의 gene annotation은 필터링, 정렬 및 다른 결과 세트와 결합 할 수 있으며, 표에서 직접 유전자를 선택할 수 있습니다.

SeedID	Source	Type	Start	End	Score	Strand	Phase	Attr-ID	Attr-Name	Attr-Parent
3R	FlyBase	gene	20114838	20117139	-	+	-	gene:FBgn0...	CG10887	gene:FBgn0...
3R	FlyBase	mRNA	20114838	20117139	-	+	-	transcript:FB...	CG10887-RA	gene:FBgn0...
3R	FlyBase	five_prime_U...	20114838	20114938	-	+	-	-	-	transcript:FB...
3R	FlyBase	exon	20114939	20114978	-	+	-	FBtr0083864...	transcript:FB...	-
3R	FlyBase	CDS	20115037	20115076	-	+	0	CDS:FBpp00...	transcript:FB...	-
3R	FlyBase	exon	20115037	20115076	-	+	-	FBtr0083864...	transcript:FB...	-
3R	FlyBase	CDS	20115037	20115076	-	+	2	CDS:FBpp00...	transcript:FB...	-
3R	FlyBase	CDS	20115133	20117080	-	+	1	CDS:FBpp00...	transcript:FB...	-
3R	FlyBase	exon	20115133	20117139	-	+	-	FBtr0083864...	transcript:FB...	-
3R	FlyBase	three_prime...	20117081	20117139	-	+	-	-	transcript:FB...	-
3R	FlyBase	gene	20121060	20122277	-	+	-	gene:FBgn0...	Gr92a	gene:FBgn0...
3R	FlyBase	mRNA	20121060	20122277	-	+	-	transcript:FB...	Gr92a-RA	gene:FBgn0...
3R	FlyBase	exon	20121060	20122061	-	+	-	FBtr0083865...	transcript:FB...	-
3R	FlyBase	CDS	20121060	20122061	-	+	0	CDS:FBpp00...	transcript:FB...	-
3R	FlyBase	CDS	20122119	20122274	-	+	0	CDS:FBpp00...	transcript:FB...	-
3R	FlyBase	exon	20122119	20122277	-	+	-	FBtr0083865...	transcript:FB...	-
3R	FlyBase	three_prime...	20122275	20122277	-	+	-	-	transcript:FB...	-
3R	FlyBase	ncRNA_gene	20139148	20139921	-	-	-	gene:FBgn0...	CR42836	gene:FBgn0...
3R	FlyBase	ncRNA	20139148	20139921	-	-	-	transcript:FB...	gene:FBgn0...	-
3R	FlyBase	exon	20139148	20139542	-	-	-	FBtr0303855...	transcript:FB...	-
3R	FlyBase	exon	20139624	20139921	-	-	-	FBtr0303856...	transcript:FB...	-
3R	FlyBase	ncRNA	20139148	20139921	-	-	-	transcript:FB...	gene:FBgn0...	-

## Statistics

다양한 통계 차트와 보고서를 통해 genome assembly 및 특성 분석 프로세스를 평가할 수 있을 뿐만 아니라, 결과의 생물학적 해석을 지원할 수 있습니다.

### Nx Plot



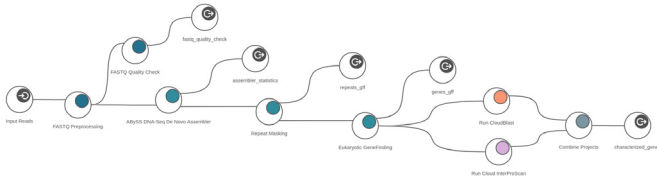
# OmicsBox with Genome Analysis

Genome의 특성을 효율적이고 사용자 친화적인 방식으로 특성화할 수 있는 Tool



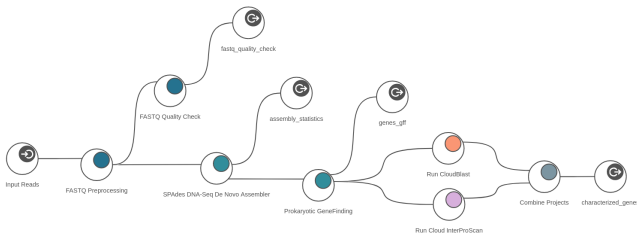
## Eukaryotic genome analysis

추가 사전 정보 없이 DNA-seq reads를 조립하여 진핵생물 종의 genome draft를 생성할 수 있고, 반복 sequence를 감지 및 마스킹하고 RNA-seq 데이터를 제공하여 유전자 예측을 개선합니다.



## Prokaryotic genome analysis

박테리아와 다른 원핵생물들의 유전자들은 빠르고 민감한 방법으로 조립되고 특징 지어질 수 있기 때문에, 결과 유전자 sequence의 functional annotation으로 진행해야 합니다.



## Long reads genome analysis

Flye로 DNA-Seq long reads를 조합하여 게놈 초안을 생성하고 short reads를 사용하여 BWA 및 Pilon으로 contig을 연마합니다. 반복적인 서열을 감지 및 마스킹하고, 유전자를 예측하고, 상동성을 찾아 조립된 서열을 완전히 특성화합니다.

