Bioinformatics in Plant Pathology

32

Aamir Khan, Sakshi Singh, and Vinay Kumar Singh

Abstract

Unprecedented success and availability of enormous next-generation sequencing data of host-pathogen in the public domain give us opportunities to understand the disease system biologically. The availability of genome data of host-pathogen in popular depository systems provides strong and proper help to retrieve, annotate, analyze and identify the functional elements for characterization at gene and genome levels for application development. The primary goal of bioinformatics is to enhance the understanding of biological processes using sequence pattern recognition, biological data mining, machine learning algorithms for biological datasets and visualization of biological data and molecules. Significant research efforts in the field include databases, software and tools development, genome analysis, anthropology, forensic genetics, sequence alignment, gene finding, genome assembly, drug design, drug discovery, protein structure alignment, protein structure prediction, gene expression analysis, microarray data analysis, protein–protein interactions and genome-wide association studies. Scientists, Paulien Hogeweg and Ben Hesper coined the term in 1970 to refer to the study of biological information processes in biotic systems. Margaret Oakley Dayhoff, the mother and father of bioinformatics compiled one of the first protein sequence databases. Elvin A. Kabat, the scientist

---

A. Khan
Division of Crop Improvement and Biotechnology, Indian Institute of Vegetable Research, Varanasi, Uttar Pradesh, India

S. Singh
Department of Molecular and Human Genetics, Institute of Science, Banaras Hindu University, Varanasi, Uttar Pradesh, India

V. K. Singh (✉)
Centre for Bioinformatics, School of Biotechnology, Institute of Science, Banaras Hindu University, Varanasi, Uttar Pradesh, India
e-mail: vinaysingh@bhu.ac.in

K. P. Singh et al. (eds.), *Emerging Trends in Plant Pathology*,
https://doi.org/10.1007/978-981-15-6275-4_32

who pioneered biological sequence analysis, developed the approach in 1970. Bioinformatics tools, techniques and databases can be used to identify potential genes, and target protein for host–pathogen interaction, drug designing and discovery and harvesting biological information from the plant genomes and their genes. Bioinformatics applications can be very beneficial in the improvement of crops and helpful for the development of designer crops.

Keywords

Bioinformatics · Plant pathology · Designer crops · Plant genomics · Next-generation sequencing

32.1 Introduction

Quality and quantity based designer crops and disease-free crops are in demand today. For that, crop improvement and protection is the first priority, in which computational biology approach for sequenced plant genomes plays a very important role and helps in crop improvement by maximizing the yield, quality-based fruits and grains production and disease resistant crops varieties (Chen and Chen 2008; King 2004; Mochida and Shinozaki 2010; Batley and Edwards 2016; Moody 2004). Development of sequence markers based on single nucleotide polymorphism and simple sequence repeat identification has now become feasible method for crop improvement. Lots of techniques, databases, tools and software have been developed to understand and analyze the biological system fully. Here standard bioinformatics techniques with specific tools and software are described.

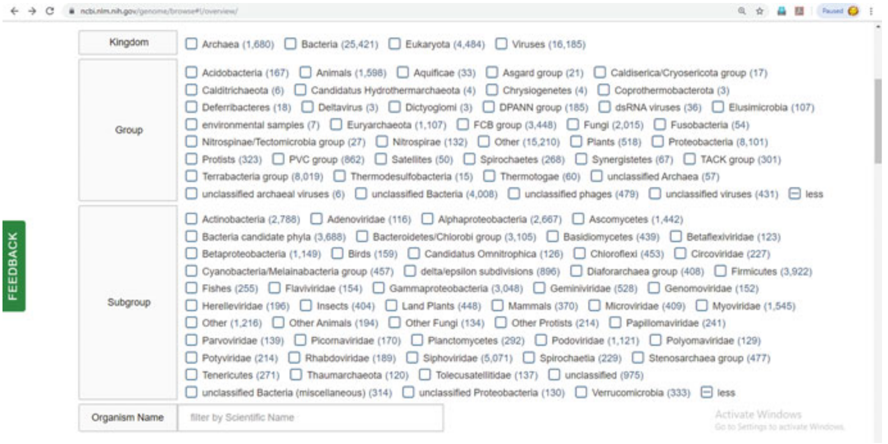
32.2 Bioinformatics Techniques

32.2.1 Comparative Analysis

A comparative analysis is a field of biological sequence analysis in which the genomic sequence features of different organisms are compared. The genomic features may include the DNA sequence, regulatory region sequence genes and gene order. The major principle of comparative analysis is that to identify the common features between homologous sequences, it will often be encoded within the DNA that is evolutionarily conserved between them or differ region which are involved in diversity (Hardison 2003; Ong et al. 2016; Gebhardt et al. 2005; Sayers et al. 2019) (Fig. 32.1).

32.2.2 Sequence Analysis

Sequence analysis is the process of subjecting a DNA, RNA or protein homologous gene (orthologous and paralogous genes) sequence to understand its evolution,



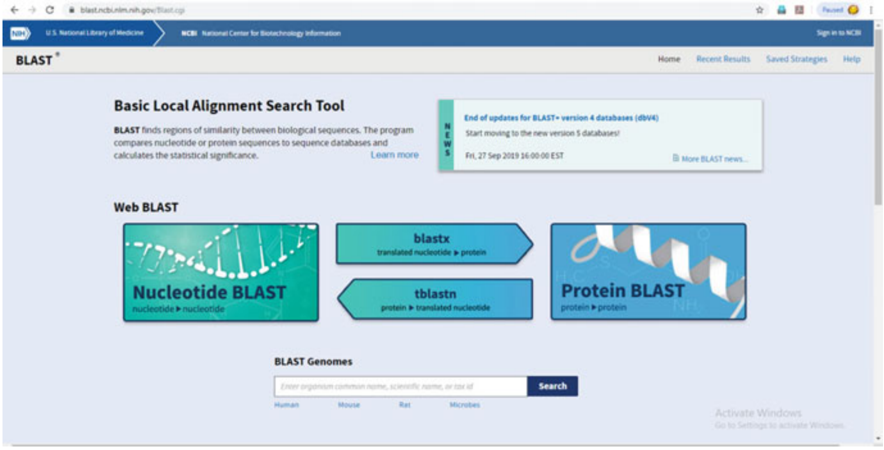
Kingdom	☐ Archaea (1,680) ☐ Bacteria (25,421) ☐ Eukaryota (4,484) ☐ Viruses (16,185)
Group	☐ Acidobacteria (167) ☐ Animals (1,598) ☐ Aquificae (33) ☐ Asgard group (21) ☐ Caldiserica/Cryosericota group (17) ☐ Calditrichaeota (6) ☐ Candidatus Hydrothermarchaeota (4) ☐ Chrysiogenetes (4) ☐ Coprothermobacterota (3) ☐ Deferribacteres (18) ☐ Deltavirus (3) ☐ Dictyoglomi (3) ☐ DPANN group (185) ☐ dsRNA viruses (36) ☐ Elusimicrobia (107) ☐ environmental samples (7) ☐ Euryarchaeota (1,107) ☐ FCB group (3,448) ☐ Fungi (2,015) ☐ Fusobacteria (54) ☐ Nitrospinae/Tectomicrobia group (27) ☐ Nitrospirae (132) ☐ Other (15,210) ☐ Plants (518) ☐ Proteobacteria (8,101) ☐ Protists (323) ☐ PVC group (862) ☐ Satellites (50) ☐ Spirochaetes (268) ☐ Synergistetes (67) ☐ TACK group (301) ☐ Terrabacteria group (8,019) ☐ Thermodesulfobacteria (15) ☐ Thermotogae (60) ☐ unclassified Archaea (57) ☐ unclassified archaeal viruses (6) ☐ unclassified Bacteria (4,008) ☐ unclassified phages (479) ☐ unclassified viruses (431) ⊟ less
Subgroup	☐ Actinobacteria (2,788) ☐ Adenoviridae (116) ☐ Alphaproteobacteria (2,667) ☐ Ascomycetes (1,442) ☐ Bacteria candidate phyla (3,688) ☐ Bacteroidetes/Chlorobi group (3,105) ☐ Basidiomycetes (439) ☐ Betaflexiviridae (123) ☐ Betaproteobacteria (1,149) ☐ Birds (159) ☐ Candidatus Omnitrophica (126) ☐ Chloroflexi (453) ☐ Circoviridae (227) ☐ Cyanobacteria/Melainabacteria group (457) ☐ delta/epsilon subdivisions (896) ☐ Diaforarchaea group (408) ☐ Firmicutes (3,922) ☐ Fishes (255) ☐ Flaviviridae (154) ☐ Gammaproteobacteria (3,048) ☐ Geminiviridae (528) ☐ Genomoviridae (152) ☐ Herelleviridae (196) ☐ Insects (404) ☐ Land Plants (448) ☐ Mammals (370) ☐ Microviridae (409) ☐ Myoviridae (1,545) ☐ Other (1,216) ☐ Other Animals (194) ☐ Other Fungi (134) ☐ Other Protists (214) ☐ Papillomaviridae (241) ☐ Parvoviridae (139) ☐ Picornaviridae (170) ☐ Planctomycetes (292) ☐ Podoviridae (1,121) ☐ Polyomaviridae (129) ☐ Potyviridae (214) ☐ Rhabdoviridae (189) ☐ Siphoviridae (5,071) ☐ Spirochaetia (229) ☐ Stenosarchaea group (477) ☐ Tenericutes (271) ☐ Thaumarchaeota (120) ☐ Tolecusatellitidae (137) ☐ unclassified (975) ☐ unclassified Bacteria (miscellaneous) (314) ☐ unclassified Proteobacteria (130) ☐ Verrucomicrobia (333) ⊟ less
Organism Name	filter by Scientific Name

Fig. 32.1 Genome availability details in the NCBI database for retrieval and comparison of sequences

function, structure or features based on sequence alignment and searches against biological sequence databases like reference genes, proteins, UniProtKB/swiss-prot, protein data bank, etc. Sequence analysis includes the comparison of common region homologous sequences in order to find similarity and dissimilarity; identification of intrinsic features of the sequence such as active sites, post-translational modification sites, gene-structures, reading frames and distributions of introns and exons and regulatory elements; identification of sequence differences and variations such as point mutations, single nucleotide variants (SNV) and single nucleotide polymorphisms (SNPs) in order to get the genetic marker, revealing the evolution and genetic diversity of sequences and organisms and identification of molecular structure from sequence alone. A basic local alignment tool is the best tool for revealing the evolutionary and genetic diversity of sequences and organisms and identification of molecular structure from sequence (Aljanabi 2001; Bolger et al. 2018; Martinez 2013; Demuth and Hahn 2009; Lyons and Freeling 2008; Altschul et al. 1990; McClure et al. 1994; Pirovano and Heringa 2008; Bawono et al. 2017) (Fig. 32.2).

32.2.3 Gene Identification

Gene hunting, gene finding or gene prediction refers to the process of identifying the regions of genomic DNA that encode genes. Gene identification is one of the first and most important steps in understanding the gene and genome of organisms once they are sequenced and available to the public domain. Gene finding is one of the key steps in genome annotation, following genome sequence assembly and the filtering of non-coding (intronic) regions and coding (exonic) regions (Alioto 2012; Wang et al. 2004; Mochida and Shinozaki 2010) (Fig. 32.3).



Basic Local Alignment Search Tool

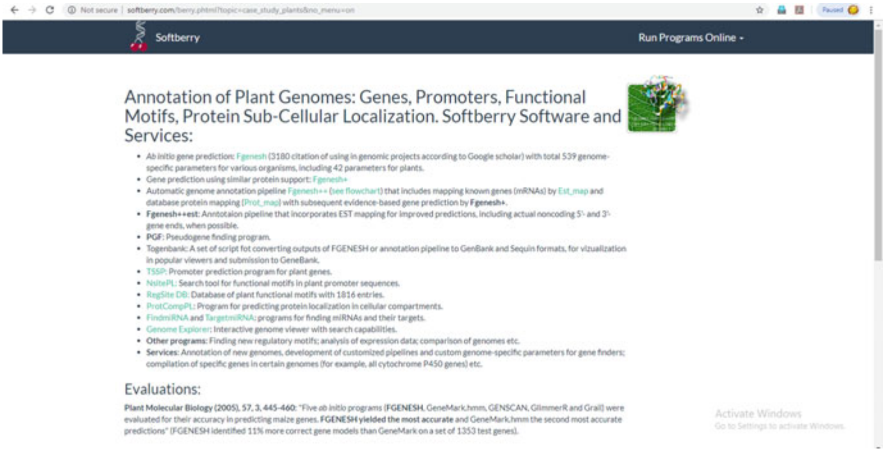
BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

End of updates for BLAST+ version 4 databases (dbV4)

Start moving to the new version 5 databases!

Fri, 27 Sep 2019 16:00:00 EST

Fig. 32.2 Basic local alignment search tool web page for sequence similarity analysis



Annotation of Plant Genomes: Genes, Promoters, Functional Motifs, Protein Sub-Cellular Localization. Softberry Software and Services:

- Ab initio gene prediction: Fgenesh (3180 citation of using in genomic projects according to Google scholar) with total 539 genome-specific parameters for various organisms, including 42 parameters for plants.
- Gene prediction using similar protein support: Fgenesh+
- Automatic genome annotation pipeline Fgenesh++ (see flowchart) that includes mapping known genes (mRNAs) by Est_map and database protein mapping (Prot_map) with subsequent evidence-based gene prediction by **Fgenesh+**.
- **Fgenesh++est**: Anntotaion pipeline that incorporates EST mapping for improved predictions, including actual noncoding 5'- and 3'-gene ends, when possible.
- **PGF**: Pseudogene finding program.
- Togenbank: A set of script fot converting outputs of FGENESH or annotation pipeline to GenBank and Sequin formats, for vizualization in popular viewers and submission to GeneBank.
- TSSP: Promoter prediction program for plant genes.
- NsitePL: Search tool for functional motifs in plant promoter sequences.
- RegSite DB: Database of plant functional motifs with 1816 entries.
- ProtCompPL: Program for predicting protein localization in cellular compartments.
- FindmiRNA and TargetmiRNA: programs for finding miRNAs and their targets.
- Genome Explorer: Interactive genome viewer with search capabilities.
- **Other programs**: Finding new regulatory motifs; analysis of expression data; comparison of genomes etc.
- **Services**: Annotation of new genomes, development of customized pipelines and custom genome-specific parameters for gene finders; compilation of specific genes in certain genomes (for example, all cytochrome P450 genes) etc.

Evaluations:

Plant Molecular Biology (2005), 57, 3, 445-460: "Five *ab initio* programs (**FGENESH**, GeneMark.hmm, GENSCAN, GlimmerR and Grail) were evaluated for their accuracy in predicting maize genes. **FGENESH yielded the most accurate** and GeneMark.hmm the second most accurate predictions" (FGENESH identified 11% more correct gene models than GeneMark on a set of 1353 test genes).

Fig. 32.3 Softberry server is a collection of software tools for genomic research focused on computational methods for high throughput biomedical data analysis

32.2.4 Phylogenetic Analysis

Phylogenetic analysis is the study of the evolutionary relationships among groups of homologous genes from organisms (e.g. species or populations). These phylogenetic relationships are discovered based on phylogenetic inference methods (distance-matrix methods: Neighbor-Joining (NJ), UPGMA (Unweighted Pair Group Method with Arithmetic mean) and WPGMA (Weighted Pair Group Method with Arithmetic mean), Fitch–Margoliash method, using outgroups, etc.; Maximum parsimony: Branch and bound, Sankoff-Morel-Cedergren algorithm, MALIGN and POY; Maximum likelihood; Bayesian inference) using sequence or morphological data. A phylogenetic tree is a branching tree diagram that represents the evolutionary relationships among selected biological organisms or species. The phylogeny inferences based on similarities and differences in their genetic or physical characteristics. Phylogenetic analyses have become central to understanding genomes, diversity, evolution and ecology (Thompson et al. 1994, 2002).

32.2.5 Protein–Protein Interaction

Protein–protein interactions (PPIs) are the physical contacts between two or more protein molecules with high specificity based on biochemical events directed by hydrophobic effect and electrostatic forces. In STRING database known interactions based on curated databases or experimentally determined, predicted interactions based on gene neighbourhood or gene fusions or gene co-occurrence and other interactions based on textmining or co-expression or protein homology (De Las Rivas and Fontanillo 2010; Kozakov et al. 2017; Szklarczyk et al. 2019) (Figs. 32.4 and 32.5).

32.2.6 Microarray Data Analysis

NCBI developed the Gene Expression Omnibus (GEO) database in 2000 for high-throughput gene expression data. Microarray data analysis is used to infer information from the data generated from DNA, RNA and protein microarray experiments; these information allows researchers to investigate the expression level of a huge number of genes of the entire organism genome in a single experiment. Gene Expression Omnibus (GEO) is a public database using MIAME (Minimum Information About a Microarray Experiment) compliant data submissions. Sequence and

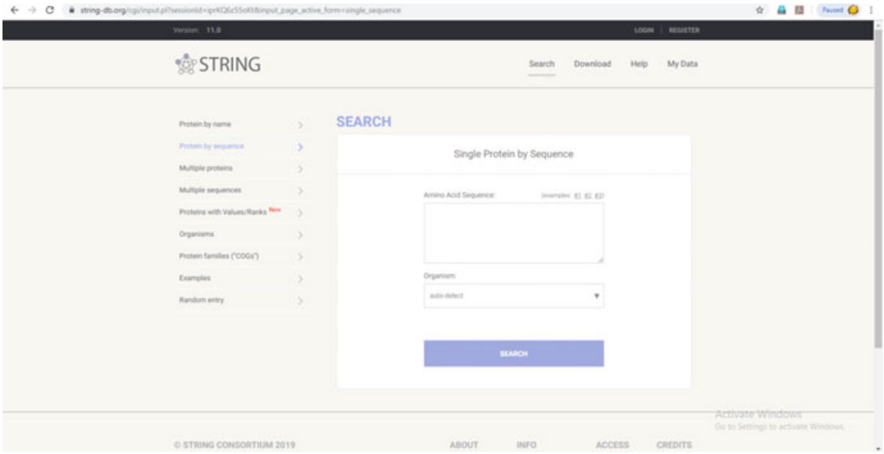


Fig. 32.4 STRING is a database for functional protein association networks

Welcome to Cluspro 2.0

Recent news: ClusPro server featured on the cover of February 2017 issue of Nature Protocols

Use Without an Account

Use the server without the benefits of your own account

--or--

Login

--or--

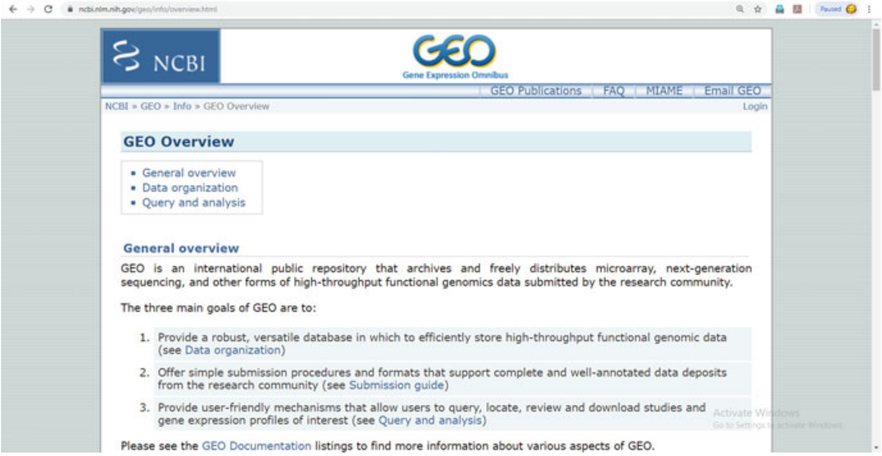
Sign up for an account

Forgot Password?
Reset Password

Forgot Username?
Retrieve Username

ClusPro should only be used for noncommercial purposes.
Structural Bioinformatics Lab and ABC Group
Boston University and Stony Brook University

Fig. 32.5 ClusPro server is a web-based server for the direct docking of two interacting proteins



GEO Overview

- General overview
- Data organization
- Query and analysis

General overview

GEO is an international public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high-throughput functional genomics data submitted by the research community.

The three main goals of GEO are to:

1. Provide a robust, versatile database in which to efficiently store high-throughput functional genomic data (see Data organization)
2. Offer simple submission procedures and formats that support complete and well-annotated data deposits from the research community (see Submission guide)
3. Provide user-friendly mechanisms that allow users to query, locate, review and download studies and gene expression profiles of interest (see Query and analysis)

Please see the GEO Documentation listings to find more information about various aspects of GEO.

Fig. 32.6 Gene Expression Omnibus (GEO) is a database repository of high throughput gene expression data and microarrays

array-based data are accepted by the repository. Techniques and tools are available to help researchers query and download experimental datasets and gene expression profiles. GEO has collected repository and it consists freely available microarray data, next-generation sequencing data, and other high-throughput functional genomics data submitted by the scientific community (Clough and Barrett 2016). Due to the complexity of data which are generated by experiments are analyzed by bioinformaticians and bio scientists with specialized softwares. GEO has developed many tools for data query, analysis and visualization that can be analyzed directly on the GEO server (Fig. 32.6).

32.2.7 Structure Prediction and Refinement

Protein structure prediction is the construction of the three-dimensional (3D) structure of a protein from its amino acid sequence. In three-dimensional structure, the 3D prediction contains folds and secondary and tertiary structures from its primary sequence. It is highly important in drug designing and in the designing of 3D novel enzymes (Krieger et al. 2003; Xiang 2006; França 2015; Cavasotto and Phatak 2009; Xu et al. 2000).

32.2.8 Molecular Docking Calculation

Molecular docking is the interaction of two or more molecules to provide a stable complex structure. Based on the binding properties of the ligand and target, it generates a three-dimensional structure complex. Molecular docking is an approach to predict the orientation of one molecule to second molecule in the bound structure, which forms a stable complex. Knowledge of the active site orientation in turn may be useful in predicting the binding strength or binding affinity between receptor-ligand molecules using scoring functions. Molecular docking is a prominent method for structure-based drug design, due to the prediction of the binding-conformation of molecular ligands to the target receptor binding site. Characterization of the active binding behaviour plays an important role in rational design of novel pesticides, herbicides, insecticides and fungicides (Ferreira et al. 2015; Guedes et al. 2014; Morris and Lim-Wilby 2008; Meng et al. 2011; de Ruyck et al. 2016; Pagadala et al. 2017; Zhao and Caflisch 2015; Kroemer 2007; Sousa et al. 2006; Jones and Willett 1995; Lybrand 1995; Goodsell et al. 1996; Gschwend et al. 1996; Trosset and Cavé 2019).

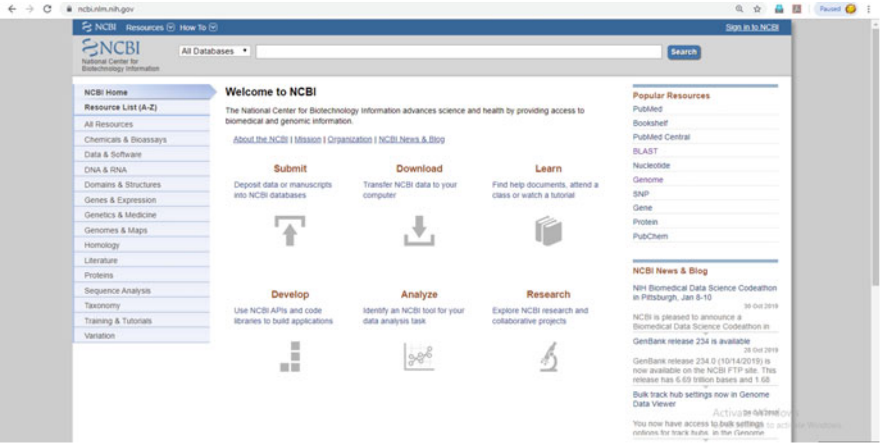
32.3 Bioinformatics Databases

Biological Data Model

Biological data model is a library of biological life sciences information and biological databases; it has a collection of computational analysis tools, literature and high-throughput experimental data. Biological database contains information from research areas including genomics, phylogenetics, proteomics, metabolomics microarray gene expression and phenomics. Information contained in biological databases includes gene structure and function, macromolecular structure, cellular and chromosomal localization and SNP and mutations in sequences and structures (Wheeler et al. 2005; Galperin and Fernández-Suárez 2012). NCBI is a data model that contains popular search engine Entrez. Entrez is NCBI's retrieval system and primary text search that integrates the PubMed and PMC database of biomedical literature with so many molecular databases including genome, gene, DNA, genetic variation, gene expression, protein sequence and structure.

32.3.1 NCBI

NCBI stands for the National Center for Biotechnology Information and is strongly associated with the National Library of Medicine (NLM) and National Institutes of Health (NIH), Bethesda, Maryland. The NCBI was founded in 1988 by Senator Claude Pepper. NCBI resources contain chemicals and bioassays data, data and software, DNA and RNA sequence data, domains and structures, genes and expression data, genetics and medicine, genomes and maps, homology data, literature, protein sequence and structure, sequence analysis, taxonomy, training and tutorials data and variation data (NCBI Resource Coordinators 2016; Wheeler et al. 2005) (Figs. 32.7, 32.8, and 32.9).



NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | NCBI News & Blog

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Develop
Use NCBI APIs and code libraries to build applications

Analyze
Identify an NCBI tool for your data analysis task

Research
Explore NCBI research and collaborative projects

Popular Resources
PubMed
Bookshelf
PubMed Central
BLAST
Nucleotide
Genome
SNP
Gene
Protein
PubChem

NCBI News & Blog

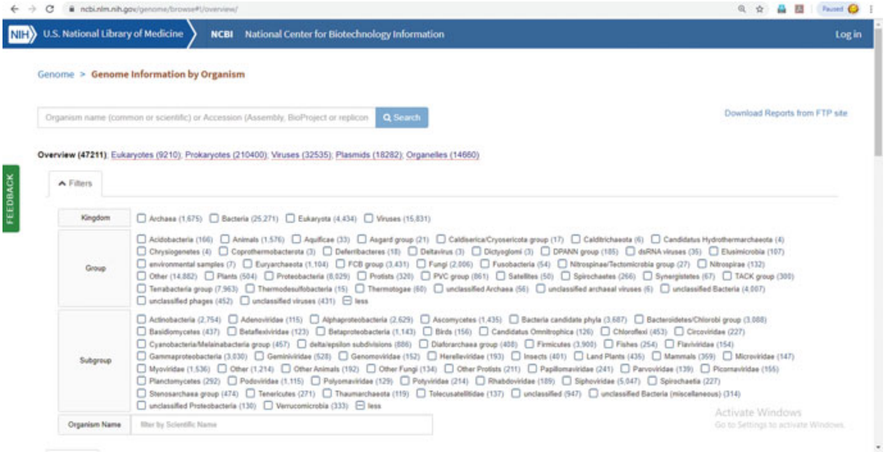
NIH Biomedical Data Science Codeathon in Pittsburgh, Jan 8-10
30 Oct 2019
NCBI is pleased to announce a Biomedical Data Science Codeathon in

GenBank release 234 is available
28 Oct 2019
GenBank release 234.0 (10/14/2019) is now available on the NCBI FTP site. This release has 6.69 trillion bases and 1.68

Bulk track hub settings now in Genome Data Viewer
You now have access to bulk settings options for track hubs in the Genome

Activate Windows
Go to Settings to activate Windows

Fig. 32.7 National Center for Biotechnology Information web page



Genome > Genome Information by Organism

Organism name (common or scientific) or Accession (Assembly, BioProject or replicon

Download Reports from FTP site

Overview (47211): Eukaryotes (9210); Prokaryotes (210400); Viruses (32535); Plasmids (18282); Organelles (14660)

Kingdom
☐ Archaea (1,675) ☐ Bacteria (25,271) ☐ Eukaryota (4,434) ☐ Viruses (15,831)

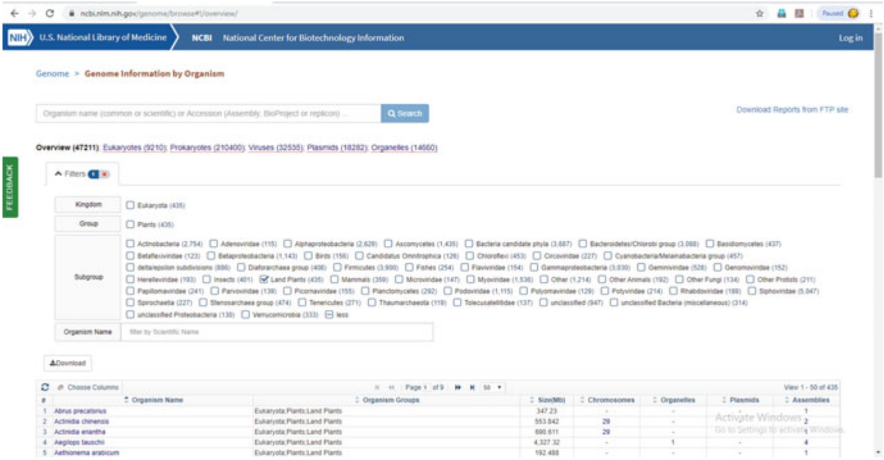
Group
☐ Acidobacteria (166) ☐ Animals (1,576) ☐ Aquificae (33) ☐ Asgard group (21) ☐ Caldiserica/Cryosericota group (17) ☐ Calditrichaeota (6) ☐ Candidatus Hydrothermarchaeota (4)
☐ Chrysiogenetes (4) ☐ Coprothermobacterota (3) ☐ Deferribacteres (18) ☐ Deltavirus (3) ☐ Dictyoglomi (3) ☐ DPANN group (185) ☐ dsRNA viruses (35) ☐ Elusimicrobia (107)
☐ environmental samples (7) ☐ Euryarchaeota (1,104) ☐ FCB group (3,431) ☐ Fungi (2,006) ☐ Fusobacteria (54) ☐ Nitrospinae/Tectomicrobia group (27) ☐ Nitrospirae (132)
☐ Other (14,882) ☐ Plants (504) ☐ Proteobacteria (8,029) ☐ Protists (320) ☐ PVC group (861) ☐ Satellites (50) ☐ Spirochaetes (266) ☐ Synergistetes (67) ☐ TACK group (300)
☐ Terrabacteria group (7,963) ☐ Thermodesulfobacteria (15) ☐ Thermotogae (60) ☐ unclassified Archaea (56) ☐ unclassified archaeal viruses (6) ☐ unclassified Bacteria (4,007)
☐ unclassified phages (452) ☐ unclassified viruses (431) ⊟ less

Subgroup
☐ Actinobacteria (2,754) ☐ Adenoviridae (115) ☐ Alphaproteobacteria (2,629) ☐ Ascomycetes (1,435) ☐ Bacteria candidate phyla (3,687) ☐ Bacteroidetes/Chlorobi group (3,088)
☐ Basidiomycetes (437) ☐ Betaflexiviridae (123) ☐ Betaproteobacteria (1,143) ☐ Birds (156) ☐ Candidatus Omnitrophica (126) ☐ Chloroflexi (453) ☐ Circoviridae (227)
☐ Cyanobacteria/Melainabacteria group (457) ☐ delta/epsilon subdivisions (886) ☐ Diaforarchaea group (408) ☐ Firmicutes (3,900) ☐ Fishes (254) ☐ Flaviviridae (154)
☐ Gammaproteobacteria (3,030) ☐ Geminiviridae (528) ☐ Genomoviridae (152) ☐ Herelleviridae (193) ☐ Insects (401) ☐ Land Plants (435) ☐ Mammals (369) ☐ Microviridae (147)
☐ Myoviridae (1,536) ☐ Other (1,214) ☐ Other Animals (192) ☐ Other Fungi (134) ☐ Other Protists (211) ☐ Papillomaviridae (241) ☐ Parvoviridae (139) ☐ Picornaviridae (155)
☐ Planctomycetes (292) ☐ Podoviridae (1,115) ☐ Polyomaviridae (129) ☐ Potyviridae (214) ☐ Rhabdoviridae (189) ☐ Siphoviridae (5,047) ☐ Spirochaetia (227)
☐ Stenosarchaea group (474) ☐ Tenericutes (271) ☐ Thaumarchaeota (119) ☐ Tolecusatellitidae (137) ☐ unclassified (947) ☐ unclassified Bacteria (miscellaneous) (314)
☐ unclassified Proteobacteria (130) ☐ Verrucomicrobia (333) ⊟ less

Organism Name
filter by Scientific Name

Activate Windows
Go to Settings to activate Windows.

Fig. 32.8 NCBI genome details page1 (The genome information can search by different kingdoms, groups, subgroups, organism name present in the NCBI database)



ncbi.nlm.nih.gov/genome/browse#!/overview/

Genome > **Genome Information by Organism**

Overview (47211): Eukaryotes (9210); Prokaryotes (210400); Viruses (32535); Plasmids (18282); Organelles (14560)

Kingdom: ☐ Eukaryota (435)

Group: ☐ Plants (435)

Subgroup:
☐ Actinobacteria (2,754) ☐ Adenoviridae (115) ☐ Alphaproteobacteria (2,629) ☐ Ascomycetes (1,435) ☐ Bacteria candidate phyla (3,687) ☐ Bacteroidetes/Chlorobi group (3,088) ☐ Basidiomycetes (437)
☐ Betaflexiviridae (123) ☐ Betaproteobacteria (1,143) ☐ Birds (156) ☐ Candidatus Omnitrophica (126) ☐ Chloroflexi (453) ☐ Circoviridae (227) ☐ Cyanobacteria/Melainabacteria group (457)
☐ delta/epsilon subdivisions (886) ☐ Diaforarchaea group (408) ☐ Firmicutes (3,900) ☐ Fishes (254) ☐ Flaviviridae (154) ☐ Gammaproteobacteria (3,030) ☐ Geminiviridae (528) ☐ Genomoviridae (152)
☐ Herelleviridae (193) ☐ Insects (401) ☑ Land Plants (435) ☐ Mammals (359) ☐ Microviridae (147) ☐ Myoviridae (1,536) ☐ Other (1,214) ☐ Other Animals (192) ☐ Other Fungi (134) ☐ Other Protists (211)
☐ Papillomaviridae (241) ☐ Parvoviridae (139) ☐ Picornaviridae (155) ☐ Planctomycetes (292) ☐ Podoviridae (1,115) ☐ Polyomaviridae (129) ☐ Polyviridae (214) ☐ Rhabdoviridae (188) ☐ Siphoviridae (5,047)
☐ Spirochaetia (227) ☐ Stenosarchaea group (474) ☐ Tenericutes (271) ☐ Thaumarchaeota (119) ☐ Tolecusatellitidae (137) ☐ unclassified (947) ☐ unclassified Bacteria (miscellaneous) (314)
☐ unclassified Proteobacteria (130) ☐ Verrucomicrobia (333) ⊟ less

Organism Name: filter by Scientific Name

Page 1 of 9 · 50 · View 1 - 50 of 435

#	Organism Name	Organism Groups	Size(Mb)	Chromosomes	Organelles	Plasmids	Assemblies
1	Abrus precatorius	Eukaryota;Plants;Land Plants	347.23	-	-	-	1
2	Actinidia chinensis	Eukaryota;Plants;Land Plants	553.842	29	-	-	2
3	Actinidia eriantha	Eukaryota;Plants;Land Plants	690.611	29	-	-	1
4	Aegilops tauschii	Eukaryota;Plants;Land Plants	4,327.32	-	1	-	4
5	Aethionema arabicum	Eukaryota;Plants;Land Plants	192.488	-	-	-	1

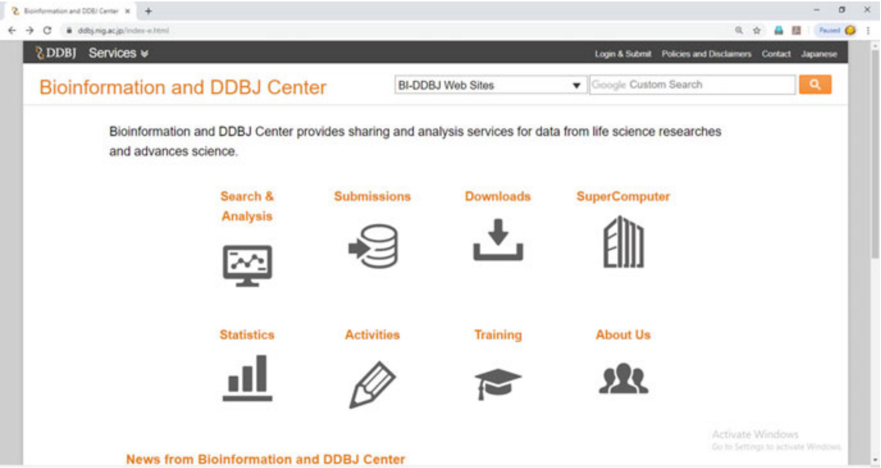
Fig. 32.9 NCBI genome details page2 (The genome information of eukaryota kingdom, plants group with their subgroups)

32.3.2 DDBJ

DDBJ (DNA Data Bank of Japan), founded in 1986, is a biological databank that mainly contains DNA sequence information. DDBJ is located at National Institute of Genetics (NIG), Shizuoka prefecture, Japan. It is also a member of INSDC (International Nucleotide Sequence Database Collaboration). The INSDC consists of a joint effort to collect and share DNA and RNA sequence data with GenBank (USA) and the European Nucleotide Archive (UK). DDBJ Sequence Read Archive (DRA), NCBI Sequence Read Archive (SRA) and EBI Sequence Read Archive (ERA) share new data and updated data on nucleotide sequences, and each of the three databases (DDBJ, NCBI and EMBL) are synchronized on a daily basis through continuous interaction between the staff at each of the collaborating organizations (Kodama et al. 2012) (Fig. 32.10).

32.3.3 EMBL

European Molecular Biology Laboratory (EMBL) is a research institution supported by 25 member states. EMBL was founded in 1974 and is a molecular biology research organization funded by public money from its member states conducted by approximately 85 independent groups. The web-based submission systems include WebIn at EMBL-EBI, Sakura ("cherry blossoms") at DDBJand BankIt at the NCBI (Madeira et al. 2019) (Fig. 32.11).



Bioinformation and DDBJ Center

Bioinformation and DDBJ Center provides sharing and analysis services for data from life science researches and advances science.

News from Bioinformation and DDBJ Center

Fig. 32.10 DNA Data Bank of Japan web homepage

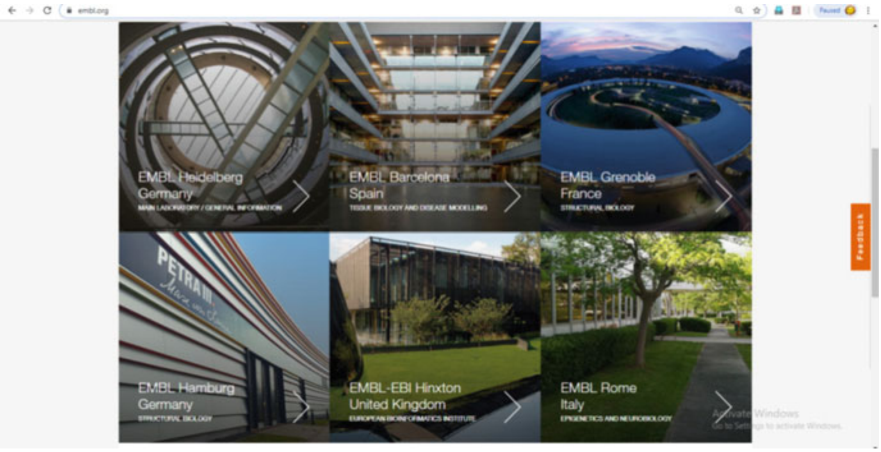
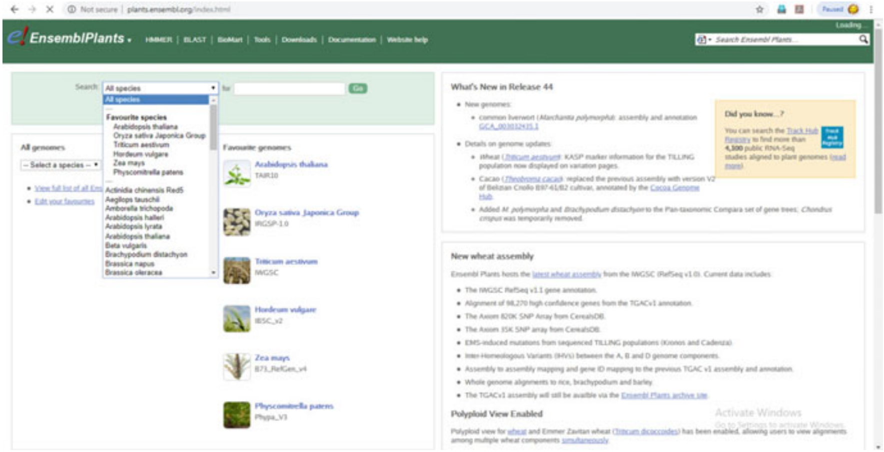


Fig. 32.11 European Molecular Biology Laboratory web page

32.3.4 Ensembl Plants

Ensembl Plants is an integrative database containing genome-scale information of plants. Ensembl Plants database includes genome sequence, gene models, polymorphic loci and functional annotation and various tools for analysis of sequence data. It contains various additional information, such as variation data, individual genotype data, linkage, population structure and phenotype data (Bolser et al. 2016, 2017) (Fig. 32.12).



What's New in Release 44

- New genomes:
  - common liverwort (*Marchantia polymorpha*): assembly and annotation GCA_003032435.1
- Details on genome updates:
  - Wheat (*Triticum aestivum*): KASP marker information for the TILLING population now displayed on variation pages.
  - Cacao (*Theobroma cacao*): replaced the previous assembly with version V2 of Belizian Criollo B97-61/B2 cultivar, annotated by the Cocoa Genome Hub.
  - Added *M. polymorpha* and *Brachypodium distachyon* to the Pan-taxonomic Compara set of gene trees; *Chondrus crispus* was temporarily removed.

Did you know...?

You can search the Track Hub Registry to find more than **4,300** public RNA-Seq studies aligned to plant genomes (read more)

New wheat assembly

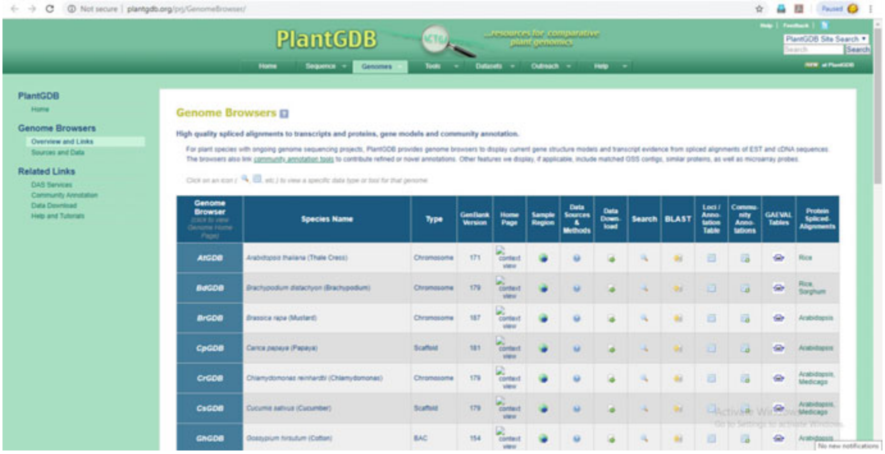
Ensembl Plants hosts the latest wheat assembly from the IWGSC (RefSeq v1.0). Current data includes:

- The IWGSC RefSeq v1.1 gene annotation.
- Alignment of 98,270 high confidence genes from the TGACv1 annotation.
- The Axiom 820K SNP Array from CerealsDB.
- The Axiom 35K SNP array from CerealsDB.
- EMS-induced mutations from sequenced TILLING populations (Kronos and Cadenza).
- Inter-Homeologous Variants (IHVs) between the A, B and D genome components.
- Assembly to assembly mapping and gene ID mapping to the previous TGAC v1 assembly and annotation.
- Whole genome alignments to rice, brachypodium and barley.
- The TGACv1 assembly will still be availble via the Ensembl Plants archive site.

Polyploid View Enabled

Polyploid view for wheat and Emmer Zavitan wheat (Triticum dicoccoides) has been enabled, allowing users to view alignments among multiple wheat components simultaneously.

Fig. 32.12 Ensembl Plants front page for genome-scale information of plant species



Genome Browsers

High quality spliced alignments to transcripts and proteins, gene models and community annotation.

For plant species with ongoing genome sequencing projects, PlantGDB provides genome browsers to display current gene structure models and transcript evidence from spliced alignments of EST and cDNA sequences. The browsers also link community annotation tools to contribute refined or novel annotations. Other features we display, if applicable, include matched GSS contigs, similar proteins, as well as microarray probes.

Click on an icon (🔍, ▢, etc.) to view a specific data type or tool for that genome.

Genome Browser (click to view Genome Home Page)	Species Name	Type	GenBank Version	Home Page	Sample Region	Data Sources & Methods	Data Down-load	Search	BLAST	Loci / Anno-tation Table	Commu-nity Anno-tations	GAEVAL Tables	Protein Spliced-Alignments
AtGDB	Arabidopsis thaliana (Thale Cress)	Chromosome	171	context view	●	●	●	🔍	●	▢	▢	●	Rice
BdGDB	Brachypodium distachyon (Brachypodium)	Chromosome	179	context view	●	●	●	🔍	●	▢	▢	●	Rice, Sorghum
BrGDB	Brassica rapa (Mustard)	Chromosome	187	context view	●	●	●	🔍	●	▢	▢	●	Arabidopsis
CpGDB	Carica papaya (Papaya)	Scaffold	181	context view	●	●	●	🔍	●	▢	▢	●	Arabidopsis
CrGDB	Chlamydomonas reinhardtii (Chlamydomonas)	Chromosome	179	context view	●	●	●	🔍	●	▢	▢	●	Arabidopsis, Medicago
CsGDB	Cucumis sativus (Cucumber)	Scaffold	179	context view	●	●	●	🔍	●	▢	▢	●	Arabidopsis, Medicago
GhGDB	Gossypium hirsutum (Cotton)	BAC	154	context view	●	●	●	🔍	●	▢	▢	●	Arabidopsis

Fig. 32.13 PlantGDB database for the comparative plant genomics information

32.3.5 PlantGDB

PlantGDB is a resource for comparative genomics and a database of molecular sequence data for plant genomes. PlantGDB contains assembled unique transcripts (PUT), genome survey sequence assemblies (GSS), genome browsers and workflow Management (Dong et al. 2004; Duvick et al. 2008) (Fig. 32.13).

32.3.6 Phytozome

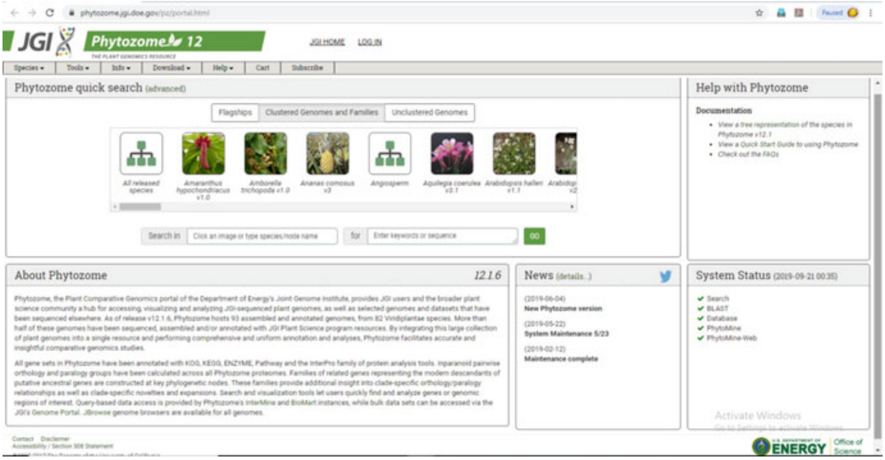
Phytozome is a comparative hub for plant genomes and gene family's data and analysis. Phytozome provides a view of genome organization, gene family, gene structure and the evolutionary history of gene at the level of sequence. It also provides access to the sequences and functional annotations of plant genomes and genes (Goodstein et al. 2012) (Fig. 32.14).

32.3.7 UNIPROT

UniProt database is a freely accessible database for protein sequence and functional annotation information, many entries being derived from different genome sequencing projects. UniProt contains a large amount of biological function of protein information derived from the literature mining. The main aim of UniProt is to provide a freely accessible resource, comprehensive and high-quality information of protein sequence and functional annotation information to scientific community (UniProt Consortium 2018) (Fig. 32.15).

32.3.8 PDB

PDB (Protein Data Bank) is a databank for the three-dimensional (3D) structural data of a large number of biological molecules, such as nucleic acids and proteins. The structural data is typically obtained by X-ray crystallography, NMR spectroscopy and cryo-electron microscopy. They are submitted by structural biologists from all around the world and are freely accessible on the net via website URLs. PDBmain



Phytozome, the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute, provides JGI users and the broader plant science community a hub for accessing, visualizing and analyzing JGI-sequenced plant genomes, as well as selected genomes and datasets that have been sequenced elsewhere. As of release v12.1.6, Phytozome hosts 93 assembled and annotated genomes, from 82 Viridiplantae species. More than half of these genomes have been sequenced, assembled and/or annotated with JGI Plant Science program resources. By integrating this large collection of plant genomes into a single resource and performing comprehensive and uniform annotation and analyses, Phytozome facilitates accurate and insightful comparative genomics studies.

All gene sets in Phytozome have been annotated with KOG, KEGG, ENZYME, Pathway and the InterPro family of protein analysis tools. Inparanoid pairwise orthology and paralogy groups have been calculated across all Phytozome proteomes. Families of related genes representing the modern descendants of putative ancestral genes are constructed at key phylogenetic nodes. These families provide additional insight into clade-specific orthology/paralogy relationships as well as clade-specific novelties and expansions. Search and visualization tools let users quickly find and analyze genes or genomic regions of interest. Query-based data access is provided by Phytozome's InterMine and BioMart instances, while bulk data sets can be accessed via the JGI's Genome Portal. JBrowse genome browsers are available for all genomes.

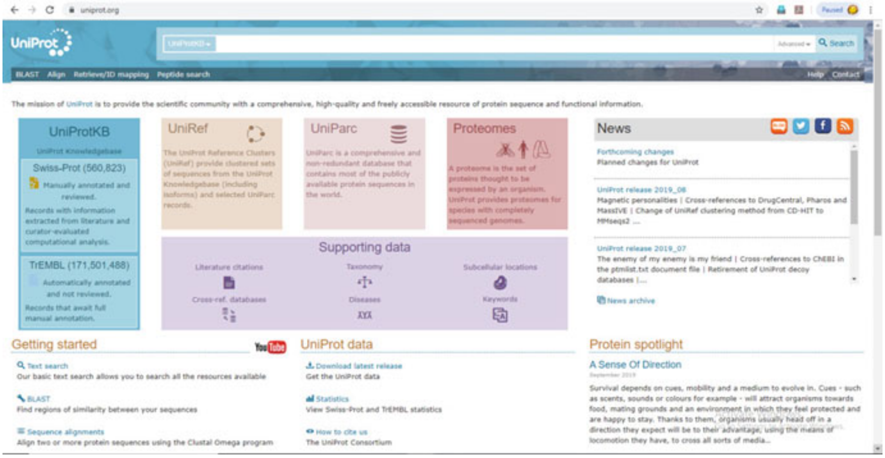
(2019-06-04)
New Phytozome version

(2019-05-22)
System Maintenance 5/23

(2019-02-12)
Maintenance complete

- ✓ Search
- ✓ BLAST
- ✓ Database
- ✓ PhytoMine
- ✓ PhytoMine-Web

Fig. 32.14 Homepage of Phytozome database



The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

Swiss-Prot (560,823): Manually annotated and reviewed. Records with information extracted from literature and curator-evaluated computational analysis.

TrEMBL (171,501,488): Automatically annotated and not reviewed. Records that await full manual annotation.

UniRef: The UniProt Reference Clusters (UniRef) provide clustered sets of sequences from the UniProt Knowledgebase (including isoforms) and selected UniParc records.

UniParc: UniParc is a comprehensive and non-redundant database that contains most of the publicly available protein sequences in the world.

Proteomes: A proteome is the set of proteins thought to be expressed by an organism. UniProt provides proteomes for species with completely sequenced genomes.

News

Forthcoming changes
Planned changes for UniProt

UniProt release 2019_08
Magnetic personalities | Cross-references to DrugCentral, Pharos and MassIVE | Change of UniRef clustering method from CD-HIT to MMseqs2 ...

UniProt release 2019_07
The enemy of my enemy is my friend | Cross-references to ChEBI in the ptmlist.txt document file | Retirement of UniProt decoy databases |...

News archive

Getting started

Text search
Our basic text search allows you to search all the resources available

BLAST
Find regions of similarity between your sequences

Sequence alignments
Align two or more protein sequences using the Clustal Omega program

UniProt data

Download latest release
Get the UniProt data

Statistics
View Swiss-Prot and TrEMBL statistics

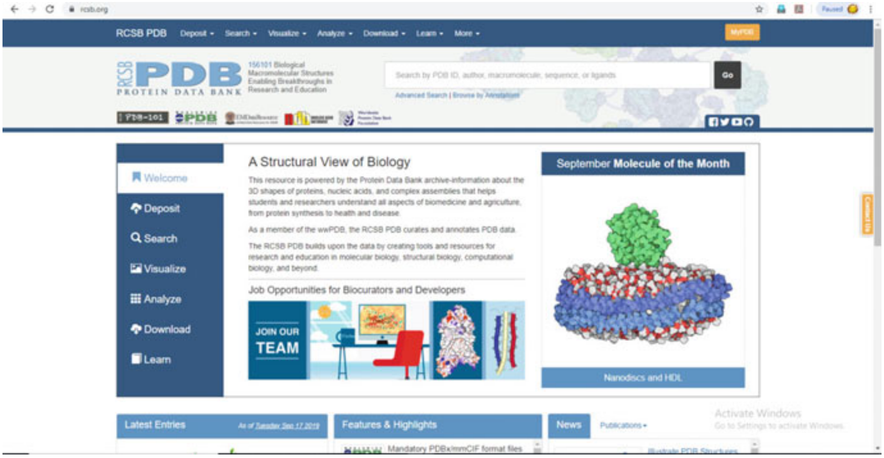
How to cite us
The UniProt Consortium

Protein spotlight

A Sense Of Direction
September 2019

Survival depends on cues, mobility and a medium to evolve in. Cues - such as scents, sounds or colours for example - will attract organisms towards food, mating grounds and an environment in which they feel protected and are happy to stay. Thanks to them, organisms usually head off in a direction they expect will be to their advantage, using the means of locomotion they have, to cross all sorts of media...

Fig. 32.15 UniProt database



A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Job Opportunities for Biocurators and Developers

Mandatory PDBx/mmCIF format files

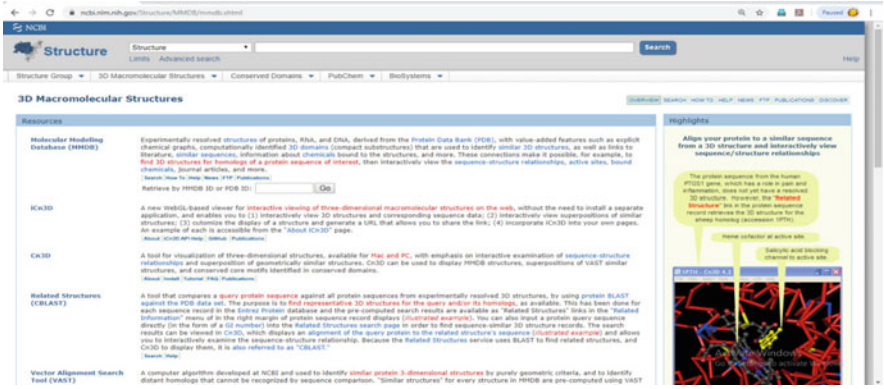
Illustrate PDB Structures

Fig. 32.16 Protein Data Bank homepage

member organizations are PDBe, PDBj, RCSB and BMRB. The PDB is overseen by an international organization called the Worldwide Protein Data Bank, wwPDB (Berman et al. 2000; Berman 2008; Laskowski et al. 1997) (Fig. 32.16).

32.3.9 MMDB

The Molecular Modeling Database (MMDB) is a three-dimensional biomolecular structure database of experimentally determined macromolecules and hosted by the National Center for Biotechnology Information (Chen et al. 2003) (Fig. 32.17).



3D Macromolecular Structures

Resources

Molecular Modeling Database (MMDB) Experimentally resolved structures of proteins, RNA, and DNA, derived from the Protein Data Bank (PDB), with value-added features such as explicit chemical graphs, computationally identified 3D domains (compact substructures) that are used to identify similar 3D structures, as well as links to literature, similar sequences, information about chemicals bound to the structures, and more. These connections make it possible, for example, to find 3D structures for homologs of a protein sequence of interest, then interactively view the sequence-structure relationships, active sites, bound chemicals, journal articles, and more.

iCn3D A new WebGL-based viewer for interactive viewing of three-dimensional macromolecular structures on the web, without the need to install a separate application, and enables you to (1) interactively view 3D structures and corresponding sequence data; (2) interactively view superpositions of similar structures; (3) cutomize the display of a structure and generate a URL that allows you to share the link; (4) incorporate iCn3D into your own pages. An example of each is accessible from the "About iCn3D" page.

Cn3D A tool for visualization of three-dimensional structures, available for Mac and PC, with emphasis on interactive examination of sequence-structure relationships and superposition of geometrically similar structures. Cn3D can be used to display MMDB structures, superpositions of VAST similar structures, and conserved core motifs identified in conserved domains.

Related Structures (CBLAST) A tool that compares a query protein sequence against all protein sequences from experimentally resolved 3D structures, by using protein BLAST against the PDB data set. The purpose is to find representative 3D structures for the query and/or its homologs, as available. This has been done for each sequence record in the Entrez Protein database and the pre-computed search results are available as "Related Structures" links in the "Related Information" menu of in the right margin of protein sequence record displays (illustrated example). You can also input a protein query sequence directly (in the form of a GI number) into the Related Structures search page in order to find sequence-similar 3D structure records. The search results can be viewed in Cn3D, which displays an alignment of the query protein to the related structure's sequence (illustrated example) and allows you to interactively examine the sequence-structure relationship. Because the Related Structures service uses BLAST to find related structures, and Cn3D to display them, it is also referred to as "CBLAST."

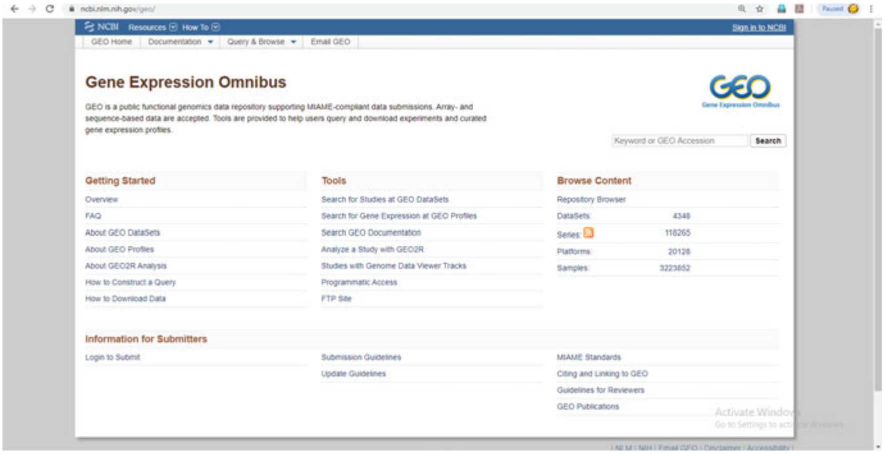
Vector Alignment Search Tool (VAST) A computer algorithm developed at NCBI and used to identify similar protein 3-dimensional structures by purely geometric criteria, and to identify distant homologs that cannot be recognized by sequence comparison. "Similar structures" for every structure in MMDB are pre-computed using VAST

Highlights

Align your protein to a similar sequence from a 3D structure and interactively view sequence/structure relationships

The protein sequence from the human PTGS1 gene, which has a role in pain and inflammation, does not yet have a resolved 3D structure. However, the "Related Structure" link in the protein sequence record retrieves the 3D structure for the sheep homolog (accession 1PTH).

Fig. 32.17 Molecular modeling database of NCBI



Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.

Getting Started
- Overview
- FAQ
- About GEO DataSets
- About GEO Profiles
- About GEO2R Analysis
- How to Construct a Query
- How to Download Data

Tools
- Search for Studies at GEO DataSets
- Search for Gene Expression at GEO Profiles
- Search GEO Documentation
- Analyze a Study with GEO2R
- Studies with Genome Data Viewer Tracks
- Programmatic Access
- FTP Site

Browse Content

Repository Browser

DataSets:	4348
Series:	118265
Platforms:	20128
Samples:	3223852

Information for Submitters
- Login to Submit
- Submission Guidelines
- Update Guidelines
- MIAME Standards
- Citing and Linking to GEO
- Guidelines for Reviewers
- GEO Publications

| NLM | NIH | Email GEO | Disclaimer | Accessibility |

Fig. 32.18 Gene Expression Omnibus database of deposited high-throughput gene expression profiling data

32.3.10 GEO

GEO (Gene Expression Omnibus) is a gene expression database that archives and freely distributes microarray datasets, next-generation sequencing analysis details and other high-throughput functional genomics datasets deposited by the research community. The main goals of GEO are to provide versatile and robust database in which researchers can efficiently store high-throughput functional genomic data, offer simple submission procedures and formats to the research community that supports complete and well-annotated data deposits and provide user-friendly mechanisms to researchers that allow users to review, query, locate and download studies and gene expression profiles of interest for query and analysis (Clough and Barrett 2016) (Fig. 32.18).

32.4 Bioinformatics Tools and Software

32.4.1 BiGGEsTS

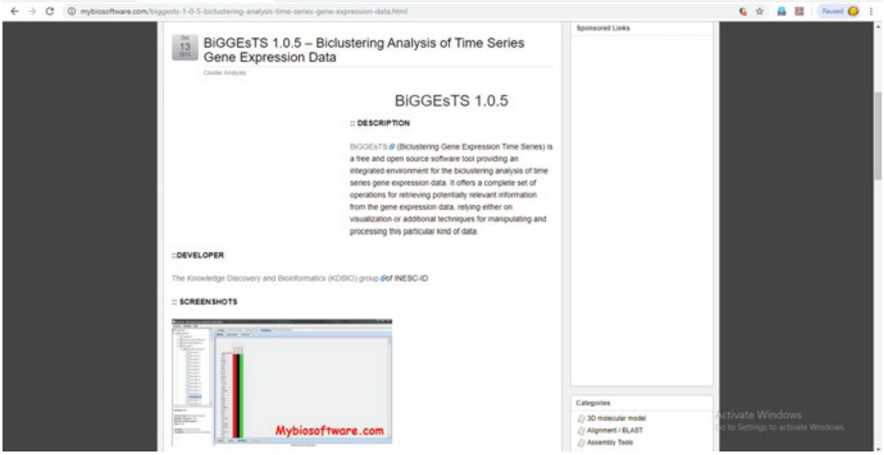
BiclusterinG Gene Expression Time Series (BiGGEsTS) is a free tool and graphical application based on bi-clustering algorithms mainly developed for analysis of gene expression time series data (Gonçalves et al. 2009) (Fig. 32.19).

32.4.2 HCE

HCE (Hierarchical Clustering Explorer) consists of hierarchical clustering algorithm to enable researchers to determine the grouping of data with informative dendrogram and colour mosaic visual feedback and dynamic query controls (Seo et al. 2006) (Fig. 32.20).

32.4.3 ClustVis

ClustVis is a web tool which allows researchers to upload their data and create Heat maps and PCA (Principal Component Analysis) plots. Data can be uploaded as a file or by pasting data to the text box (Metsalu and Vilo 2015) (Fig. 32.21).



mybiosoftware.com/biggests-1-0-5-biclustering-analysis-time-series-gene-expression-data.html

BiGGEsTS 1.0.5 – Biclustering Analysis of Time Series Gene Expression Data

BiGGEsTS 1.0.5

:: DESCRIPTION

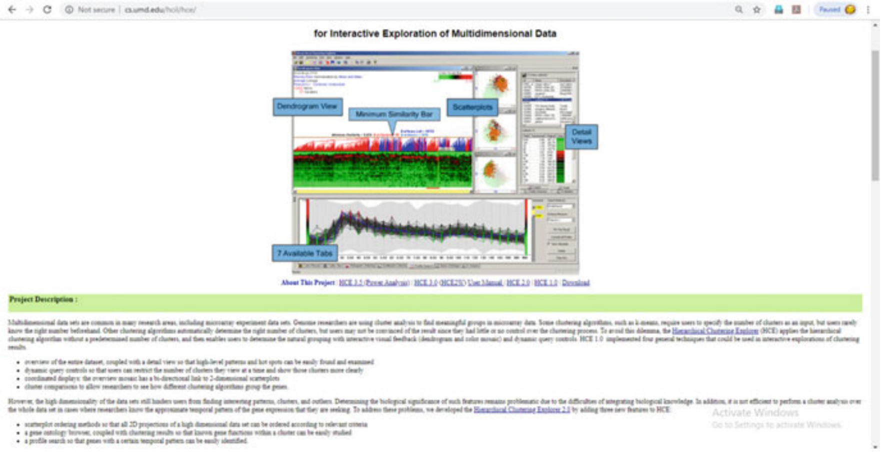
BiGGEsTS (Biclustering Gene Expression Time Series) is a free and open source software tool providing an integrated environment for the biclustering analysis of time series gene expression data. It offers a complete set of operations for retrieving potentially relevant information from the gene expression data, relying either on visualization or additional techniques for manipulating and processing this particular kind of data.

::DEVELOPER

The Knowledge Discovery and Bioinformatics (KDBIO) group of INESC-ID

:: SCREENSHOTS

Fig. 32.19 BiclusterinG Gene Expression Time Series



for Interactive Exploration of Multidimensional Data

About This Project | HCE 3.5 (Power Analysis) | HCE 3.0 (HCE2W) User Manual | HCE 2.0 | HCE 1.0 | Download

Project Description :

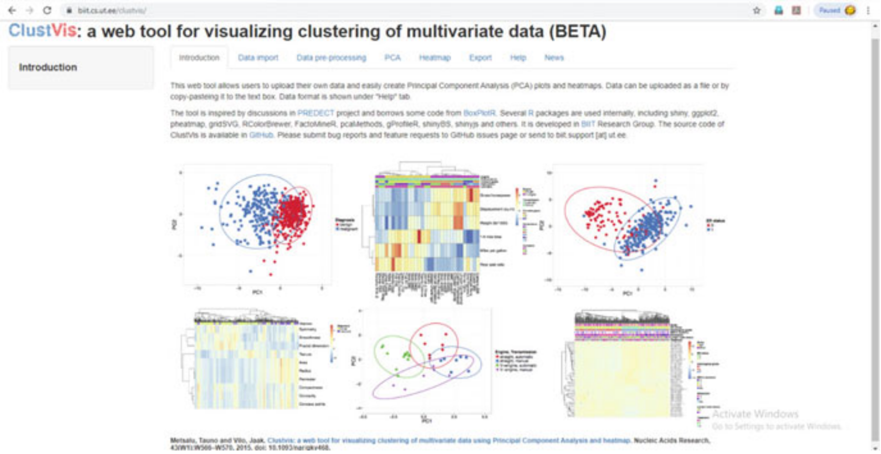
Multidimensional data sets are common in many research areas, including microarray experiment data sets. Genome researchers are using cluster analysis to find meaningful groups in microarray data. Some clustering algorithms, such as k-means, require users to specify the number of clusters as an input, but users rarely know the right number beforehand. Other clustering algorithms automatically determine the right number of clusters, but users may not be convinced of the result since they had little or no control over the clustering process. To avoid this dilemma, the Hierarchical Clustering Explorer (HCE) applies the hierarchical clustering algorithm without a predetermined number of clusters, and then enables users to determine the natural grouping with interactive visual feedback (dendrogram and color mosaic) and dynamic query controls. HCE 1.0 implemented four general techniques that could be used in interactive explorations of clustering results.

- overview of the entire dataset, coupled with a detail view so that high-level patterns and hot spots can be easily found and examined
- dynamic query controls so that users can restrict the number of clusters they view at a time and show those clusters more clearly
- coordinated displays: the overview mosaic has a bi-directional link to 2-dimensional scatterplots
- cluster comparisons to allow researchers to see how different clustering algorithms group the genes.

However, the high dimensionality of the data sets still hinders users from finding interesting patterns, clusters, and outliers. Determining the biological significance of such features remains problematic due to the difficulties of integrating biological knowledge. In addition, it is not efficient to perform a cluster analysis over the whole data set in cases where researchers know the approximate temporal pattern of the gene expression that they are seeking. To address these problems, we developed the Hierarchical Clustering Explorer 2.0 by adding three new features to HCE:

- scatterplot ordering methods so that all 2D projections of a high dimensional data set can be ordered according to relevant criteria
- a gene ontology browser, coupled with clustering results so that known gene functions within a cluster can be easily studied
- a profile search so that genes with a certain temporal pattern can be easily identified.

Fig. 32.20 Hierarchical Clustering Explorer



ClustVis: a web tool for visualizing clustering of multivariate data (BETA)

This web tool allows users to upload their own data and easily create Principal Component Analysis (PCA) plots and heatmaps. Data can be uploaded as a file or by copy-pasteing it to the text box. Data format is shown under "Help" tab.

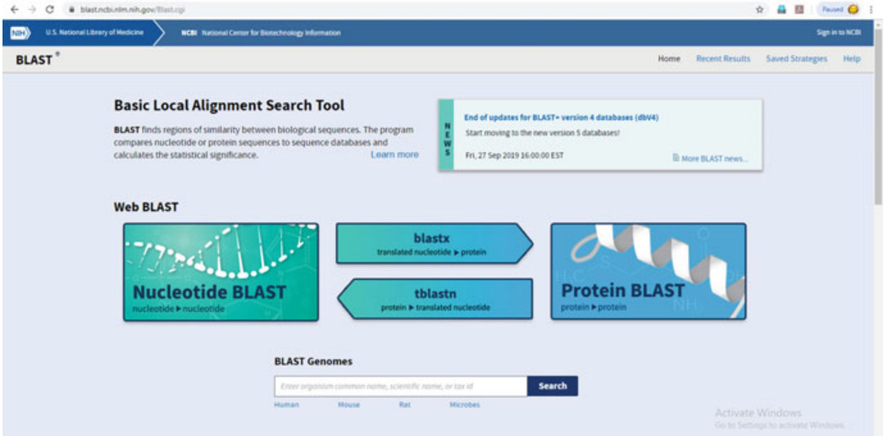
The tool is inspired by discussions in PREDECT project and borrows some code from BoxPlotR. Several R packages are used internally, including shiny, ggplot2, pheatmap, gridSVG, RColorBrewer, FactoMineR, pcaMethods, gProfileR, shinyBS, shinyjs and others. It is developed in BIIT Research Group. The source code of ClustVis is available in GitHub. Please submit bug reports and feature requests to GitHub issues page or send to biit.support [at] ut.ee.

Metsalu, Tauno and Vilo, Jaak. Clustvis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. Nucleic Acids Research, 43(W1):W566–W570, 2015. doi: 10.1093/nar/gkv468.

Fig. 32.21 ClustVis web tool

32.4.4 BLAST

BLAST (Basic Local Alignment Search Tool) finds regions of similarity and dissimilarity between sequences. The BLAST programme compares nucleotide or protein sequences to sequence databases and calculates identity with statistical significance (Altschul et al. 1990; Mount 2007) (Fig. 32.22).



Basic Local Alignment Search Tool

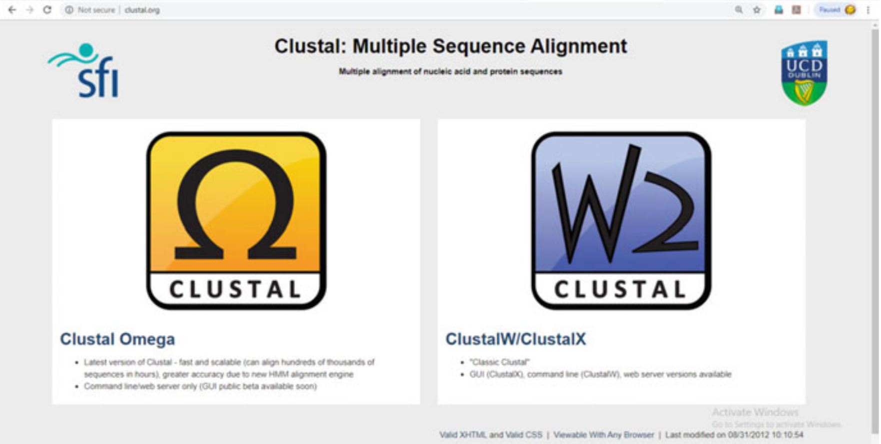
BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

End of updates for BLAST+ version 4 databases (dbV4)

Start moving to the new version 5 databases!

Fri, 27 Sep 2019 16:00:00 EST

Fig. 32.22 Basic Local Alignment Search Tool



Clustal: Multiple Sequence Alignment

Multiple alignment of nucleic acid and protein sequences

Clustal Omega

- Latest version of Clustal - fast and scalable (can align hundreds of thousands of sequences in hours), greater accuracy due to new HMM alignment engine
- Command line/web server only (GUI public beta available soon)

ClustalW/ClustalX

- "Classic Clustal"
- GUI (ClustalX), command line (ClustalW), web server versions available

Valid XHTML and Valid CSS | Viewable With Any Browser | Last modified on 08/31/2012 10:10:54

Fig. 32.23 Clustal series homepage

32.4.5 Clustal

Clustal omega, *Clustalw* and Clustalx (Clustal series) are widely used programmes for multiple sequence alignment (Higgins et al. 1996; Chenna et al. 2003; Sievers and Higgins 2014) (Fig. 32.23).

32.4.6 Bioedit

BioEdit is a free sequence alignment editor for editing and manipulation of sequence alignment data (Tippmann 2004) (Fig. 32.24).

Copyright © 1997-2005
Tom Hall
Ibis Therapeutics
Carlsbad, CA 92008

Biological sequence alignment editor for Win95/98/NT/2K/XP

BioEdit is a biological sequence alignment editor written for Windows 95/98/NT/2000/XP. An intuitive multiple document interface with convenient features makes alignment and manipulation of sequences relatively easy on your desktop computer. Several sequence manipulation and analysis options and links to external anaylsis programs facilitate a working environment which allows you to view and manipulate sequences with simple point-and-click operations.

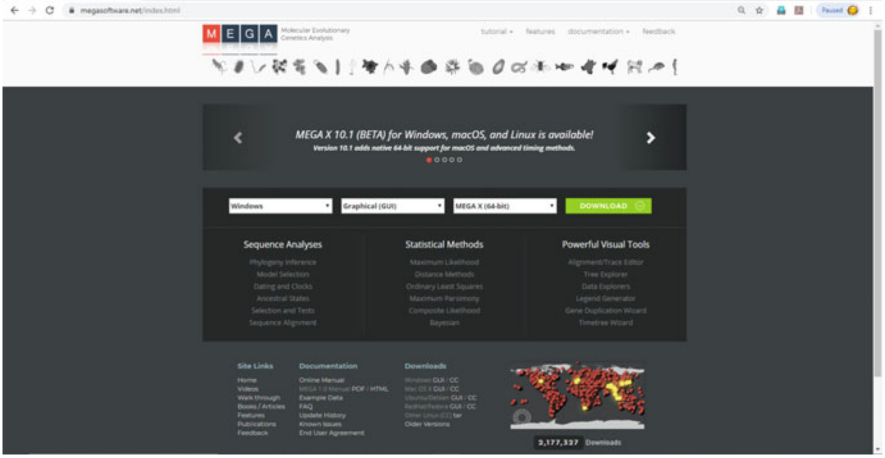
BioEdit's features include: New version is WinXP compatible

- Several modes of hand alignment
- Automated ClustalW alignment
- Automated Blast searches (local and WWW)
- Plasmid drawing and annotation
- Accessory application configuration
- Restriction mapping
- RNA comparative analysis tools
- Graphical matrix data viewing tools
- Shaded alignment figures
- Translation-based nucleic acid alignment
- ABI trace viewing, editing and printing
- Customizable ... other features

Note: Although BioEdit was recently updated, it is no longer being reliably maintained, and the documentation is out of date and no longer maintained. It is being updated slowly, but there is no guaranteed finish date. Until documentation is complete, play with the menus and see what happens, or email with a question.

Note: If you have trouble launching BioEdit on Windows NT, try replacing BioEdit.exe with this version

Fig. 32.24 BioEdit is a biological sequence alignment editor tool



MEGA X 10.1 (BETA) for Windows, macOS, and Linux is available!
Version 10.1 adds native 64-bit support for macOS and advanced timing methods.

Fig. 32.25 Molecular evolutionary genetic analysis

32.4.7 MEGA

MEGA is a tool for manual and automatic sequence alignment, phylogenetic tree preparation, estimating rates of molecular evolution, web-based database mining and testing evolutionary hypotheses (Kumar et al. 2018) (Fig. 32.25).

32.4.8 Figtree

Figtree is a graphical viewer of phylogenetic tree visualization and for producing publication-ready figures of phylogenetic trees (Rambaut 2012) (Fig. 32.26).

32.4.9 Circos

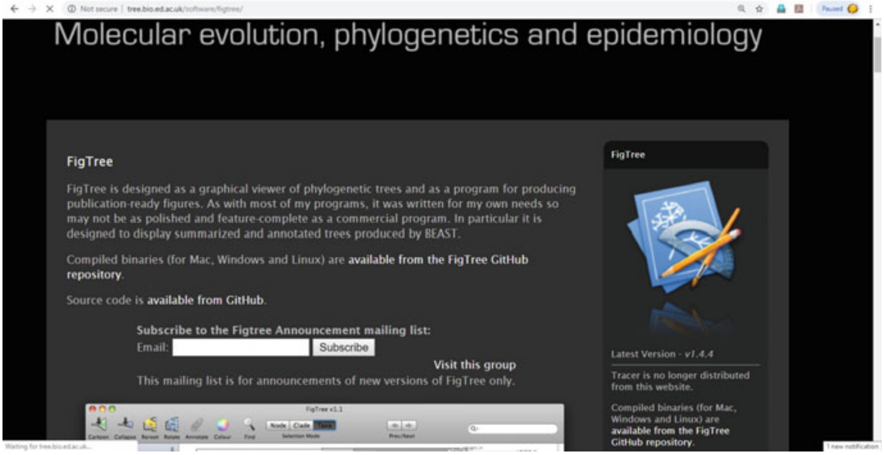
Circos server is basically for identification and analysis of similarities and dissimilarity/differences generated from gene and genome comparisons (Krzywinski et al. 2009) (Fig. 32.27).

32.4.10 Prosite

PROSITE server is protein database that consists of protein families, functional domains and functional signature sites and amino acid profiles and patterns in sequence (Sigrist et al. 2002) (Fig. 32.28).

32.4.11 CDD

Conserved Domain Database (CDD) is a protein database that consists of well-annotated multiple sequence alignments as position-specific score matrices (PSSMs) for identification of conserved domains via RPS-BLAST. CDD includes NCBI-curated functional domains based on 3D-structure information to define domain boundaries and provide functional insights into sequence/structure/function



Molecular evolution, phylogenetics and epidemiology

FigTree

FigTree is designed as a graphical viewer of phylogenetic trees and as a program for producing publication-ready figures. As with most of my programs, it was written for my own needs so may not be as polished and feature-complete as a commercial program. In particular it is designed to display summarized and annotated trees produced by BEAST.

Compiled binaries (for Mac, Windows and Linux) are **available from the FigTree GitHub repository**.

Source code is **available from GitHub**.

Subscribe to the Figtree Announcement mailing list:

Email:

Visit this group

This mailing list is for announcements of new versions of FigTree only.

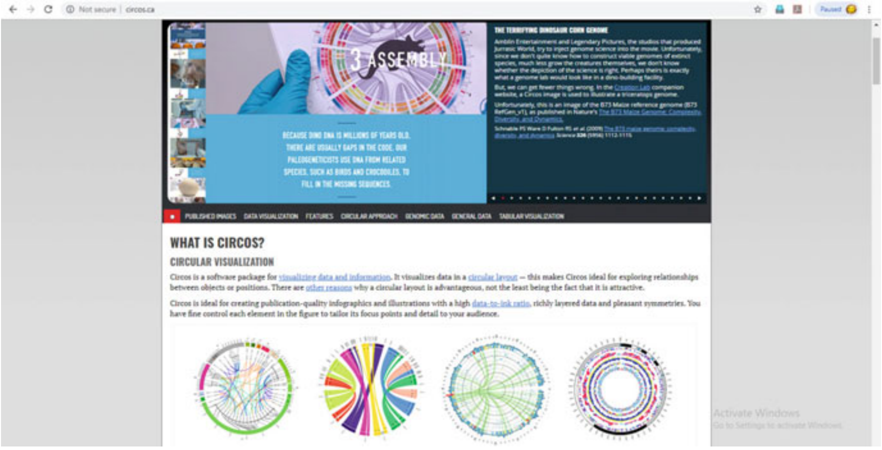
FigTree

Latest Version - v1.4.4

Tracer is no longer distributed from this website.

Compiled binaries (for Mac, Windows and Linux) are **available from the FigTree GitHub repository**.

Fig. 32.26 FigTree server



BECAUSE DINO DNA IS MILLIONS OF YEARS OLD, THERE ARE USUALLY GAPS IN THE CODE. OUR PALEOGENETICISTS USE DNA FROM RELATED SPECIES, SUCH AS BIRDS AND CROCODILES, TO FILL IN THE MISSING SEQUENCES.

THE TERRIFYING DINOSAUR CORN GENOME

Amblin Entertainment and Legendary Pictures, the studios that produced Jurassic World, try to inject genome science into the movie. Unfortunately, since we don't quite know how to construct viable genomes of extinct species, much less grow the creatures themselves, we don't know whether the depiction of the science is right. Perhaps theirs is exactly what a genome lab would look like in a dino-building facility.

But, we can get fewer things wrong. In the Creation Lab companion website, a Circos image is used to illustrate a triceratops genome.

Unfortunately, this is an image of the B73 Maize reference genome (B73 RefGen_v1), as published in Nature's The B73 Maize Genome: Complexity, Diversity and Dynamics.

Schnable PS Ware D Fulton RS et al. (2009) The B73 maize genome: complexity, diversity and dynamics Science 326 (5956) 1112-1115.

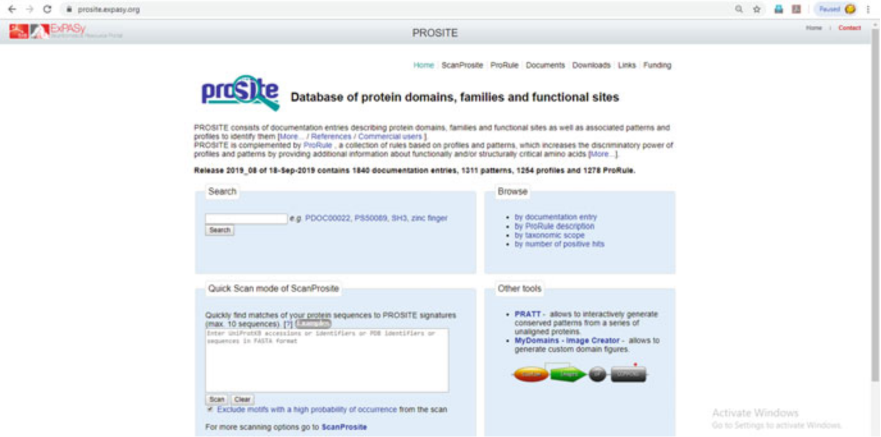
WHAT IS CIRCOS?

CIRCULAR VISUALIZATION

Circos is a software package for visualizing data and information. It visualizes data in a circular layout — this makes Circos ideal for exploring relationships between objects or positions. There are other reasons why a circular layout is advantageous, not the least being the fact that it is attractive.

Circos is ideal for creating publication-quality infographics and illustrations with a high data-to-ink ratio, richly layered data and pleasant symmetries. You have fine control each element in the figure to tailor its focus points and detail to your audience.

Fig. 32.27 Circos server



Database of protein domains, families and functional sites

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [More... / References / Commercial users].
PROSITE is complemented by ProRule , a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [More...].

Release 2019_08 of 18-Sep-2019 contains 1840 documentation entries, 1311 patterns, 1254 profiles and 1278 ProRule.

e.g. PDOC00022, PS50089, SH3, zinc finger

- by documentation entry
- by ProRule description
- by taxonomic scope
- by number of positive hits

Quickly find matches of your protein sequences to PROSITE signatures (max. 10 sequences). [?]

Enter UniProtKB accessions or identifiers or PDB identifiers or sequences in FASTA format

☑ Exclude motifs with a high probability of occurrence from the scan

For more scanning options go to **ScanProsite**

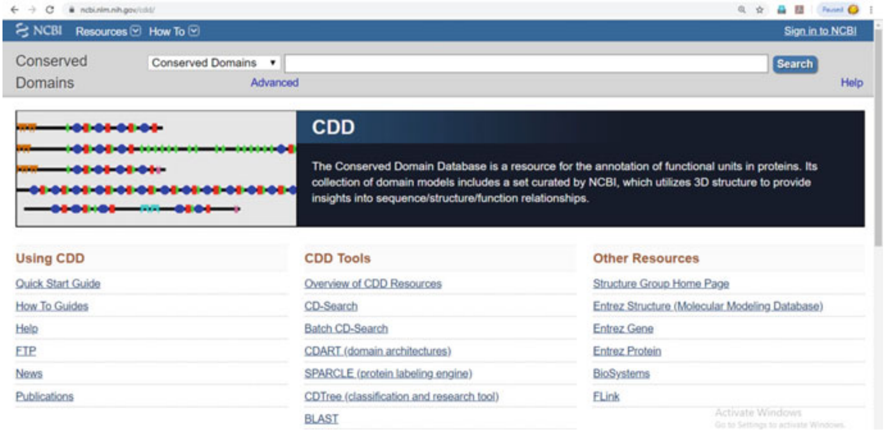
- **PRATT** - allows to interactively generate conserved patterns from a series of unaligned proteins.
- **MyDomains - Image Creator** - allows to generate custom domain figures.

Fig. 32.28 PROSITE server

relationships, using Pfam, SMART, COG, PRK and TIGRFAMs databases (Marchler-Bauer et al. 2017) (Fig. 32.29).

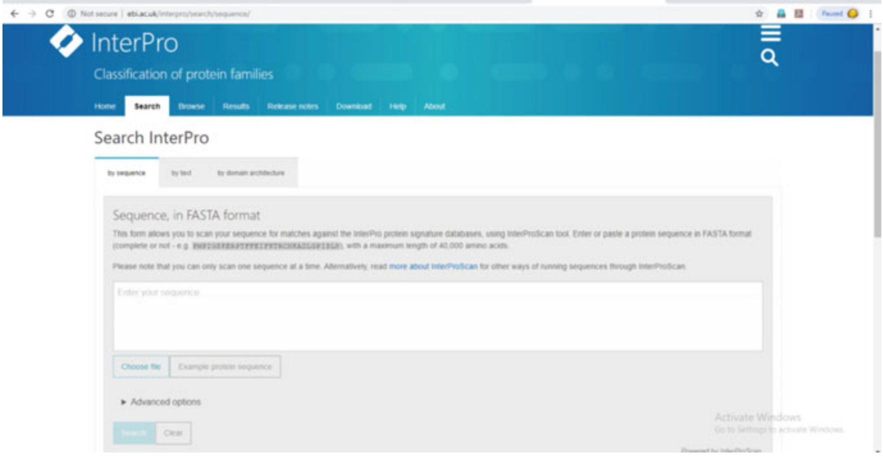
32.4.12 Interproscan

InterProScan is a server to annotate protein families and domains automatically. InterPro provides functional signature analysis of proteins by classifying them into families, domains and important sites (Mitchell et al. 2019) (Fig. 32.30).



The Conserved Domain Database is a resource for the annotation of functional units in proteins. Its collection of domain models includes a set curated by NCBI, which utilizes 3D structure to provide insights into sequence/structure/function relationships.

Fig. 32.29 Conserved Domain Database



This form allows you to scan your sequence for matches against the InterPro protein signature databases, using InterProScan tool. Enter or paste a protein sequence in FASTA format (complete or not - e.g. [illegible]), with a maximum length of 40,000 amino acids.

Please note that you can only scan one sequence at a time. Alternatively, read more about InterProScan for other ways of running sequences through InterProScan.

Fig. 32.30 InterProScan server

32.4.13 EasyModeller

EasyModeller is a graphical user interface programme used for homology modeling for predicting models of protein tertiary structures (Kuntal et al. 2010) (Fig. 32.31).

32.4.14 RAMPAGE/PROCHECK

PROCHECK server checks the stereochemical quality of a protein structure model; it produces Ramachandran plot to analyze the overall and residue-by-residue geometry (Laskowski et al. 2017; Lovell et al. 2003) (Figs. 32.32 and 32.33).

MODELLER GUI

A GRAPHICAL USER INTERFACE TO MODELLER FOR HOMOLOGY MODELLING OF PROTEIN SEQUENCES.

MONDAY, JULY 23, 2012

EasyModeller 4.0 : A new GUI to MODELLER

EasyModeller 4.0 introduces a fresh new GUI for Homology Modelling using MODELLER in the backend and available for both Windows and Linux platform. This version has several new features integrating all the goodies of EasyModeller 3.0 which was only available for LINUX.

The highlighted features of this version are :

1. Tab based logical Modelling steps with extensive error handling
2. Allows to load unlimited number of templates
3. A colorful alignment viewer and also with an inbuilt alignment editor
4. On the fly MODELLER code editing (generated MODELLER scripts can be edited as per user's need and run from within the tool). This feature would be most useful for advanced MODELLER users.
5. Inbuilt DOPE profile Viewer, Ramachandran Plot viewer, Loop Modelling, Basic model optimization and dynamics for a selected Model.

DOWNLOAD Link : Click here

CLICK HERE TO JOIN THE FORUM

EasyModeller Discussion Forum

YOU ARE VISITOR NUMBER

13128

RECOMMEND EASYMODELLER

WHICH OS + MODELLER + PYTHON WORKS BEST FOR YOU ? THIS SURVEY WILL HELP MANY USERS SO PLEASE TAKE A BIT OF YOUR TIME TO FILL IT.

Fig. 32.31 EasyModeller

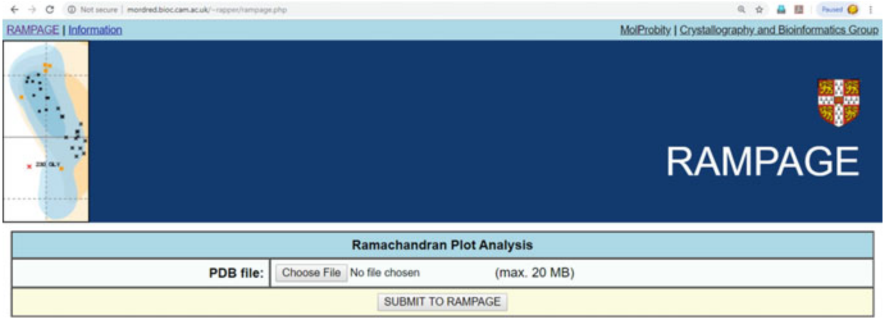


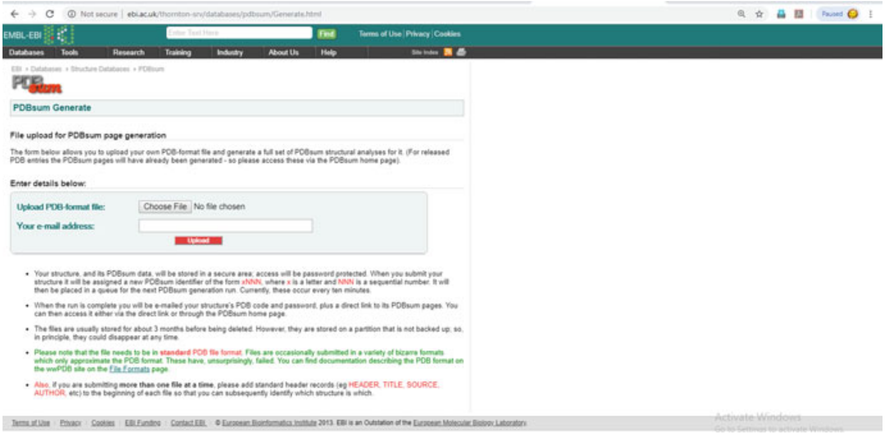
Fig. 32.32 RAMPAGE server

32.4.15 VERIFY3D

VERIFY3D server is used for determination of an atomic model (3D) with its amino acid sequence, by assigning a structural class based on alpha, beta, loop, polar, non-polar, etc. location and comparing the results to template structures (Eisenberg et al. 1997) (Fig. 32.34).

32.4.16 YASARA

YASARA (Yet Another Scientific Artificial Reality Application) is a computer programme for molecular vizualization, modeling and docking (Krieger and Vriend 2014) (Fig. 32.35).



PDBsum Generate

File upload for PDBsum page generation

The form below allows you to upload your own PDB-format file and generate a full set of PDBsum structural analyses for it. (For released PDB entries the PDBsum pages will have already been generated - so please access these via the PDBsum home page).

Enter details below:

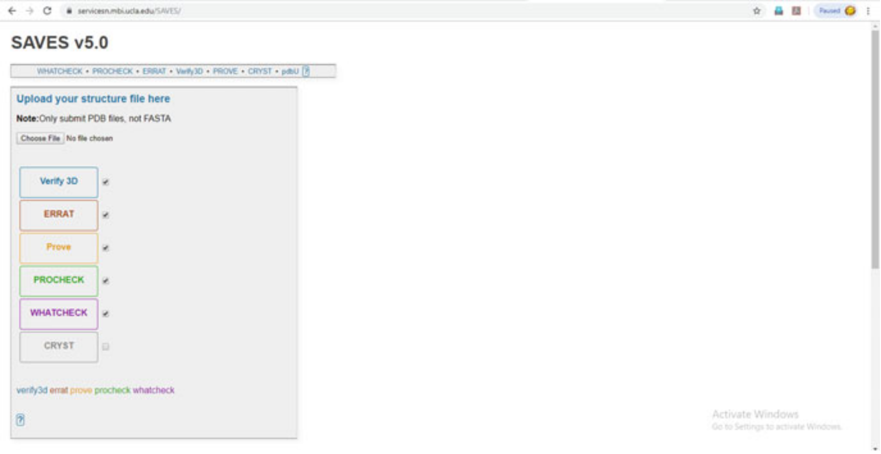
Upload PDB-format file: No file chosen

Your e-mail address:

- Your structure, and its PDBsum data, will be stored in a secure area; access will be password protected. When you submit your structure it will be assigned a new PDBsum identifier of the form xNNN, where x is a letter and NNN is a sequential number. It will then be placed in a queue for the next PDBsum generation run. Currently, these occur every ten minutes.
- When the run is complete you will be e-mailed your structure's PDB code and password, plus a direct link to its PDBsum pages. You can then access it either via the direct link or through the PDBsum home page.
- The files are usually stored for about 3 months before being deleted. However, they are stored on a partition that is not backed up; so, in principle, they could disappear at any time.
- Please note that the file needs to be in **standard** PDB file format. Files are occasionally submitted in a variety of bizarre formats which only approximate the PDB format. These have, unsurprisingly, failed. You can find documentation describing the PDB format on the wwPDB site on the File Formats page.
- Also, if you are submitting **more than one file at a time**, please add standard header records (eg HEADER, TITLE, SOURCE, AUTHOR, etc) to the beginning of each file so that you can subsequently identify which structure is which.

© European Bioinformatics Institute 2013. EBI is an Outstation of the European Molecular Biology Laboratory.

Fig. 32.33 PDBSum



Note: Only submit PDB files, not FASTA

No file chosen

Fig. 32.34 SAVES server

32.4.17 BIOVIA Discovery Studio 2019

BIOVIA Discovery Studio contains BIOVIA Pipeline Pilot used for simulations, macromolecule design and analysis, antibody modeling, structure-based design, pharmacophore and ligand-based design, QSAR, ADMET and predictive toxicology, X-ray and visualization (Fig. 32.36).

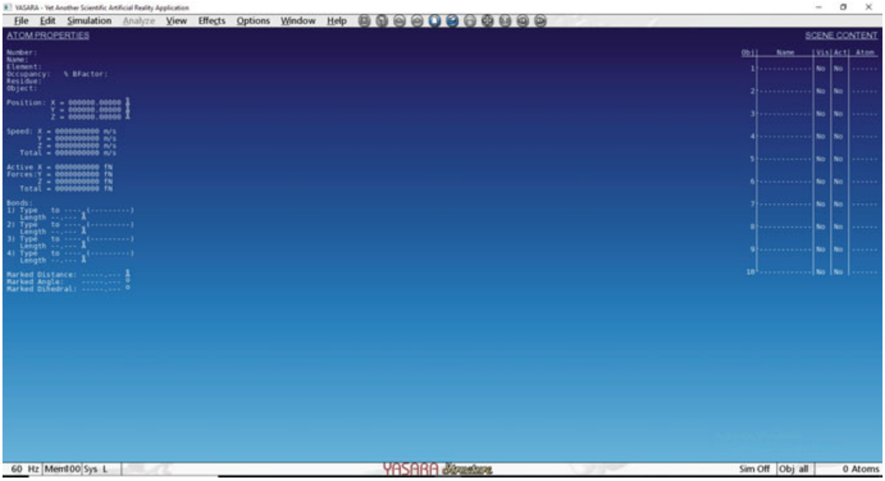


Fig. 32.35 Yet Another Scientific Artificial Reality Application

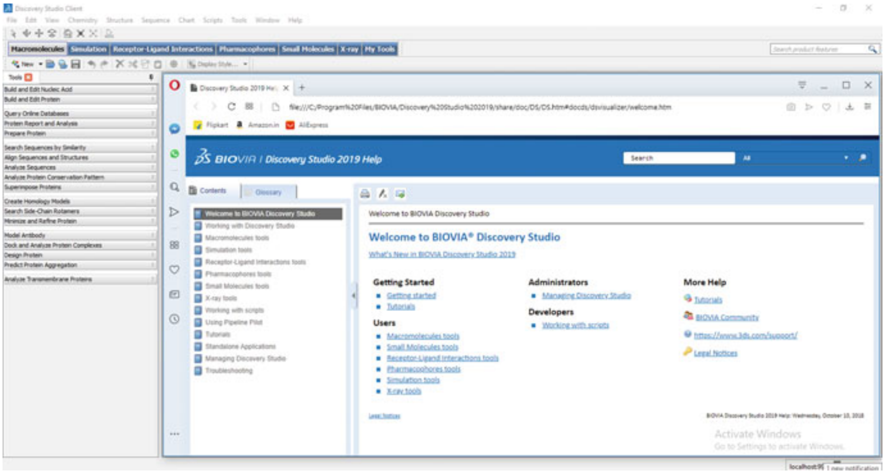
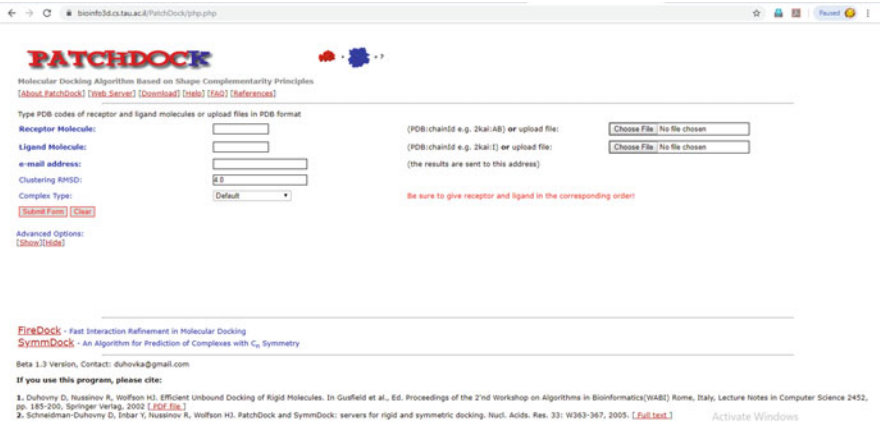


Fig. 32.36 BIOVIA Discovery Studio

32.4.18 Patchdock

The PatchDock server performs protein–protein docking and generates protein-small molecule complexes (Schneidman-Duhovny et al. 2005) (Fig. 32.37).



Molecular Docking Algorithm Based on Shape Complementarity Principles

[About PatchDock] [Web Server] [Download] [Help] [FAQ] [References]

Type PDB codes of receptor and ligand molecules or upload files in PDB format

Receptor Molecule: (PDB:chainId e.g. 2kai:AB) **or** upload file: Choose File No file chosen

Ligand Molecule: (PDB:chainId e.g. 2kai:I) **or** upload file: Choose File No file chosen

e-mail address: (the results are sent to this address)

Clustering RMSD: 4.0

Complex Type: Default

Be sure to give receptor and ligand in the corresponding order!

Submit Form | Clear

Advanced Options:
[Show][Hide]

FireDock - Fast Interaction Refinement in Molecular Docking
SymmDock - An Algorithm for Prediction of Complexes with C_n Symmetry

Beta 1.3 Version, Contact: duhovka@gmail.com

If you use this program, please cite:

1. Duhovny D, Nussinov R, Wolfson HJ. Efficient Unbound Docking of Rigid Molecules. In Gusfield et al., Ed. Proceedings of the 2'nd Workshop on Algorithms in Bioinformatics(WABI) Rome, Italy, Lecture Notes in Computer Science 2452, pp. 185-200, Springer Verlag, 2002 [PDF file]
2. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucl. Acids. Res. 33: W363-367, 2005. [Full text]

Fig. 32.37 PatchDock server

Hex Server

Docking - step 1 of 2

Receptor PDB File Choose File No file chosen

Ligand PDB File Choose File No file chosen

Email Address (Optional)

Correlation Type Shape Only

Calculation Device GPU

Search Order 25

reset | next

Could not connect to database - the compute cluster is probably off-line

Help Examples More Information

Hex KBDOCK

Fig. 32.38 Hex server

32.4.19 Hex

Hex tool/server is a graphics programme for docking calculation and visualizing docking modes of pairs of protein and DNA molecules. Hex is also useful for calculation of protein-ligand docking; it can superpose molecules (Macindoe et al. 2010) (Fig. 32.38).

32.5 Plant and Pathogen Genomics

Five main types of pathogenic organisms that cause plant diseases are viruses, bacteria, fungi, protozoa and worms/nematodes, which can lead from damage to death. The genome availability of plants and pathogens gives us opportunities to understand the bio systems and disease mechanisms (Tables 32.1, 32.2, and 32.3).

Table 32.1 List of important plant diseases with their causing organism, in which most of pathogen genomes are available in the NCBI database

Disease	Causing organism (pathogen)
Bacterial leaf blight	*Pseudomonas syringae subsp. syringae*
Aster yellows	*Phytoplasma*
Bacterial wilt	*Erwinia tracheiphila*
Bacterial blight	*Xanthomonas campestris, Xanthomonas axonopodis, Pseudomonas syringae*
Crown gall	*Agrobacterium tumefaciens*
Bacterial soft rot	*Erwinia, Pectobacterium and Pseudomonas*
Scab	*Venturia inaequalis, Streptomyces scabies*
Anthracnose	*Colletotrichum*
Black knot	*Dibotryon morbosum* or *Apiosporina morbosa*
Blight	*Cryphonectria parasitica*, *Cochliobolus heterostrophus*, *Colletotrichum capsici*
Chestnut blight	*Cryphonectria parasitica*
Late blight	*Phytophthora infestans*
Canker	*Sirococcusclavigignenti-juglandacearum*, *Seiridiumcardinale*, *Gibberellabaccata*, *Diplodiaquercina*, *Leptosphaeria coniothyrium, Cryptosporella umbrina*, *Colletotrichum coccodes*
Clubroot	*Plasmodiophora brassicae*
Damping-off	*Pythium*
Dutch elm disease	*Claviceps purpurea*
Fusarium wilt Panama disease	*Fusarium oxysporum*
Leaf blister	*Taphrina caerulescens*
Downy mildew	*Pseudoperonospora cubensis*
Powdery mildew	*Podosphaera xanthii*, *Erysiphe cichoracearum*
Oak wilt	*Ceratocystis fagacearum*
Rot	*Oomycota*
Basal rot	*Botrytis, Fusarium, and Penicillium*
Graymold rot	*Botrytis cinerea*

(continued)

Table 32.1 (continued)

Disease	Causing organism (pathogen)
Rust	*Phragmidium spp.*
Blister rust	*Cronartium ribicola*
Cedar-apple rust	*Gymnosporangium juniperi-virginianae*
Coffee rust	*Hemileia vastatrix*
Scab	*Venturia inaequalis*
Smut	*Sporisorium scitamineum*
Bunt	*Tilletia tritici*
Corn smut	*Ustilago maydis*
Sooty mold	*Cladosporium and Alternaria*
Verticillium wilt	*Verticillium*
Curly top	(*Becurtovirus, Begomovirus, Capulavirus, Curtovirus, Eragrovirus, Grablovirus, Mastrevirus, Topocuvirus, Turncurtovirus*)
Mosaic	*Tobacco mosaic virus*
Psorosis	*Citrus psorosisophiovirus*
Spotted wilt	*Tomato spotted wilt virus*
Root-knot nematodes	*Meloidogyne*
Witchweed	*Pratylenchus*

Table 32.2 List of important plant pathogen genome details

Organism name	Organism groups	Size(Mb)	Assemblies
Abutilon Brazil virus	Viruses; Geminiviridae	0.005271	1
Abutilon golden mosaic virus	Viruses; Geminiviridae	0.002629	1
Abutilon mosaic Bolivia virus	Viruses; Geminiviridae	0.005399	1
Abutilon mosaic Brazil virus	Viruses; Geminiviridae	0.005282	1
Abutilon mosaic virus	Viruses; Geminiviridae	0.005217	1
African cassava mosaic Burkina Faso virus	Viruses; Geminiviridae	0.00277	1
African cassava mosaic virus	Viruses; Geminiviridae	0.005503	1
Ageratum enation virus	Viruses; Geminiviridae	0.00276	3
Ageratum leaf curl Cameroon virus	Viruses; Geminiviridae	0.002792	1
Ageratum leaf curl virus – [G52]	Viruses; Geminiviridae	0.002735	1
Ageratum yellow vein China virus – OX1	Viruses; Geminiviridae	0.002739	1
Ageratum yellow vein Hualian virus	Viruses; Geminiviridae	0.002756	2
Ageratum yellow vein Sri Lanka virus	Viruses; Geminiviridae	0.002748	1
Ageratum yellow vein virus	Viruses; Geminiviridae	0.002768	12
Agrobacterium tumefaciens	Bacteria; Proteobacteria; Alphaproteobacteria	7.2733	69
Alfalfa leaf curl virus	Viruses; Geminiviridae	0.002745	1
Allamanda leaf curl virus	Viruses; Geminiviridae	0.002755	1
Allamanda leaf mottle distortion virus	Viruses; Geminiviridae	0.005462	1
Alternanthera yellow vein virus	Viruses; Geminiviridae	0.002745	3
Alternaria alternata	Eukaryota; Fungi; Ascomycetes	32.9908	6
Alternaria arborescens	Eukaryota; Fungi; Ascomycetes	33.9434	5
Alternaria atra	Eukaryota; Fungi; Ascomycetes	35.1121	1
Alternaria brassicae	Eukaryota; Fungi; Ascomycetes	34.1411	1
Alternaria brassicicola	Eukaryota; Fungi; Ascomycetes	31.0365	2
Alternaria consortialis	Eukaryota; Fungi; Ascomycetes	34.2409	1
Alternaria gaisen	Eukaryota; Fungi; Ascomycetes	34.3469	1
Alternaria gansuensis	Eukaryota; Fungi; Ascomycetes	75.0519	1
Alternaria solani	Eukaryota; Fungi; Ascomycetes	32.7791	2
Alternaria sp. MG1	Eukaryota; Fungi; Ascomycetes	34.6956	1
Alternaria tenuissima	Eukaryota; Fungi; Ascomycetes	35.7042	7

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Andrographis yellow vein leaf curl virus	Viruses; Geminiviridae	0.002754	1
Aster yellow witches'-broom phytoplasma	Bacteria; Terrabacteriagroup; Tenericutes	0.72397	1
Asystasia mosaic Madagascar virus	Viruses; Geminiviridae	0.005404	1
Axonopus compressus streak virus	Viruses; Geminiviridae	0.002858	1
Bean calico mosaic virus	Viruses; Geminiviridae	0.005175	1
Bean chlorosis virus	Viruses; Geminiviridae	0.005279	1
Bean dwarf mosaic virus	Viruses; Geminiviridae	0.005191	1
Bean golden mosaic virus	Viruses; Geminiviridae	0.005197	2
Bean golden yellow mosaic virus	Viruses; Geminiviridae	0.005255	2
Bean leaf crumple virus	Viruses; Geminiviridae	0.002598	1
Bean leaf curl Madagascar virus	Viruses; Geminiviridae	0.002754	1
Bean white chlorosis mosaic virus	Viruses; Geminiviridae	0.005163	1
Bean yellow dwarf virus	Viruses; Geminiviridae	0.002561	1
Bean yellow dwarf virus	Viruses; Geminiviridae	0.002561	1
Bean yellow mosaic Mexico virus	Viruses; Geminiviridae	0.002641	1
Beet curly top Iran virus	Viruses; Geminiviridae	0.002859	4
Beet curly top virus	Viruses; Geminiviridae	0.002994	8
Bhendi yellow vein Bhubhaneswar virus	Viruses; Geminiviridae	0.002757	1
Bhendi yellow vein Delhi virus [2004: New Delhi]	Viruses; Geminiviridae	0.002751	1
Bhendi yellow vein Haryana virus	Viruses; Geminiviridae	0.00274	1
Bhendi yellow vein India virus [India: Dharwad OYDWR2:2006]	Viruses; Geminiviridae	0.002739	1
Bhendi yellow vein mosaic virus	Viruses; Geminiviridae	0.002747	6
Bipolaris maydis	Eukaryota; Fungi; Ascomycetes	32.9292	2
Bitter gourd yellow vein virus	Viruses; Geminiviridae	0.005453	1
Blainvillea yellow spot virus	Viruses; Geminiviridae	0.0053	1
Blechum interveinal chlorosis virus	Viruses; Geminiviridae	0.005285	1
Boerhavia yellow spot virus	Viruses; Geminiviridae	0.002621	1
Botrytis cinerea	Eukaryota; Fungi; Ascomycetes	42.6301	4
'Brassica napus' phytoplasma	Bacteria; Terrabacteria group; Tenericutes	0.743598	1
Bretziellafagacearum	Eukaryota; Fungi; Ascomycetes	26.782	1
Bromuscatharticus striate mosaic virus	Viruses; Geminiviridae	0.002797	1
Cabbage leaf curl Jamaica virus	Viruses; Geminiviridae	0.005069	1
Cabbage leaf curl virus	Viruses; Geminiviridae	0.005096	1
Candidatus Phytoplasma	Bacteria; Terrabacteria group; Tenericutes	0.687137	1

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Candidatus Phytoplasma aurantifolia	Bacteria; Terrabacteria group; Tenericutes	0.474669	1
Candidatus Phytoplasma australiense	Bacteria; Terrabacteria group; Tenericutes	0.959779	2
Candidatus Phytoplasma mali	Bacteria; Terrabacteria group; Tenericutes	0.601943	1
Candidatus Phytoplasma oryzae	Bacteria; Terrabacteria group; Tenericutes	0.533195	2
Candidatus Phytoplasma phoenicium	Bacteria; Terrabacteria group; Tenericutes	0.541091	2
Candidatus Phytoplasma pini	Bacteria; Terrabacteria group; Tenericutes	0.474136	1
Candidatus Phytoplasma pruni	Bacteria; Terrabacteria group; Tenericutes	0.598511	1
Candidatus Phytoplasma solani	Bacteria; Terrabacteria group; Tenericutes	0.821322	3
Candidatus Phytoplasma ziziphi	Bacteria; Terrabacteria group; Tenericutes	0.750803	1
Capraria yellow spot Yucatan virus	Viruses; Geminiviridae	0.005208	1
Cassava mosaic Madagascar virus	Viruses; Geminiviridae	0.00551	1
'Catharanthus roseus' aster yellows phytoplasma	Bacteria; Terrabacteria group; Tenericutes	0.603949	1
Catharanthus yellow mosaic virus	Viruses; Geminiviridae	0.002752	1
Centrosema yellow spot virus	Viruses; Geminiviridae	0.002675	1
Chayote yellow mosaic virus	Viruses; Geminiviridae	0.002787	1
Chenopodium leaf curl virus	Viruses; Geminiviridae	0.002626	1
Chickpea chlorosis Australia virus	Viruses; Geminiviridae	0.002572	1
Chickpea chlorosis virus	Viruses; Geminiviridae	0.002603	3
Chickpea chlorosis virus-A	Viruses; Geminiviridae	0.002582	1
Chickpea chlorotic dwarf virus	Viruses; Geminiviridae	0.002587	5
Chickpea redleaf virus	Viruses; Geminiviridae	0.002605	1
Chickpea yellow dwarf virus	Viruses; Geminiviridae	0.002547	1
Chickpea yellows virus	Viruses; Geminiviridae	0.002557	1
Chilli leaf curl Ahmedabad virus-India [India/Ahmedabad/2014]	Viruses; Geminiviridae	0.002744	1
Chilli leaf curl India virus	Viruses; Geminiviridae	0.002755	1
Chilli leaf curl Kanpur virus	Viruses; Geminiviridae	0.002754	1
Chilli leaf curl Vellanad virus	Viruses; Geminiviridae	0.002788	1
Chilli leaf curl virus	Viruses; Geminiviridae	0.002858	7
Chino del tomate Amazonas virus	Viruses; Geminiviridae	0.002615	1
Chino del tomate virus	Viruses; Geminiviridae	0.005213	3
Chloris striate mosaic virus	Viruses; Geminiviridae	0.00275	1
'Chrysanthemum coronarium' phytoplasma	Bacteria; Terrabacteria group; Tenericutes	0.739592	1

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Chrysanthemum yellows phytoplasma	Bacteria; Terrabacteria group; Tenericutes	0.659699	1
Citrus psorosis virus	Viruses; Aspiviridae	0.011278	1
Cladosporium cladosporioides	Eukaryota; Fungi; Ascomycetes	33.2257	1
Cladosporium phlei	Eukaryota; Fungi; Ascomycetes	32.816	1
Cladosporium sp. SL-16	Eukaryota; Fungi; Ascomycetes	35.8569	1
Cladosporium sphaerospermum	Eukaryota; Fungi; Ascomycetes	26.8942	1
Claviceps purpurea	Eukaryota; Fungi; Ascomycetes	32.0914	3
Cleome golden mosaic virus	Viruses; Geminiviridae	0.002566	1
Cleome leaf crumple virus	Viruses; Geminiviridae	0.005386	1
Clerodendron golden mosaic virus	Viruses; Geminiviridae	0.005524	1
Clerodendron yellow mosaic virus	Viruses; Geminiviridae	0.00276	1
Clerodendrum golden mosaic China virus	Viruses; Geminiviridae	0.005515	2
Clerodendrum golden mosaic Jiangsu virus	Viruses; Geminiviridae	0.002753	1
Cnidoscolus mosaic leaf deformation virus	Viruses; Geminiviridae	0.005254	1
Coccinia mosaic Tamil Nadu virus	Viruses; Geminiviridae	0.00544	1
Colletotrichum acutatum	Eukaryota; Fungi; Ascomycetes	52.1291	2
Colletotrichum chlorophyti	Eukaryota; Fungi; Ascomycetes	52.387	1
Colletotrichum coccodes	Eukaryota; Fungi; Ascomycetes	50.122	2
Colletotrichum falcatum	Eukaryota; Fungi; Ascomycetes	48.1864	1
Colletotrichum fioriniae	Eukaryota; Fungi; Ascomycetes	50.1509	3
Colletotrichum fructicola	Eukaryota; Fungi; Ascomycetes	55.9157	3
Colletotrichum gloeosporioides	Eukaryota; Fungi; Ascomycetes	61.9165	6
Colletotrichum godetiae	Eukaryota; Fungi; Ascomycetes	35.0343	1
Colletotrichum graminicola	Eukaryota; Fungi; Ascomycetes	51.6443	2
Colletotrichum higginsianum	Eukaryota; Fungi; Ascomycetes	50.7161	3
Colletotrichum incanum	Eukaryota; Fungi; Ascomycetes	53.2546	2

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Colletotrichum lentis	Eukaryota; Fungi; Ascomycetes	56.1001	1
Colletotrichum lindemuthianum	Eukaryota; Fungi; Ascomycetes	99.1667	2
Colletotrichum musae	Eukaryota; Fungi; Ascomycetes	49.1188	1
Colletotrichum nymphaeae	Eukaryota; Fungi; Ascomycetes	49.9563	1
Colletotrichum orbiculare	Eukaryota; Fungi; Ascomycetes	89.7483	1
Colletotrichum orchidophilum	Eukaryota; Fungi; Ascomycetes	48.5565	1
Colletotrichum salicis	Eukaryota; Fungi; Ascomycetes	48.3734	1
Colletotrichum sansevieriae	Eukaryota; Fungi; Ascomycetes	51.2013	1
Colletotrichum shisoi	Eukaryota; Fungi; Ascomycetes	69.6677	1
Colletotrichum siamense	Eukaryota; Fungi; Ascomycetes	55.9616	1
Colletotrichum sidae	Eukaryota; Fungi; Ascomycetes	86.8278	1
Colletotrichum simmondsii	Eukaryota; Fungi; Ascomycetes	50.4742	1
Colletotrichum sp. JS-367	Eukaryota; Fungi; Ascomycetes	87.1965	1
Colletotrichum spinosum	Eukaryota; Fungi; Ascomycetes	82.7349	1
Colletotrichum sublineola	Eukaryota; Fungi; Ascomycetes	64.8486	2
Colletotrichum tanaceti	Eukaryota; Fungi; Ascomycetes	57.9125	1
Colletotrichum tofieldiae	Eukaryota; Fungi; Ascomycetes	52.7196	5
Colletotrichum trifolii	Eukaryota; Fungi; Ascomycetes	109.66	1
Colletotrichum truncatum	Eukaryota; Fungi; Ascomycetes	57.9128	2
Common bean mottle virus	Viruses; Geminiviridae	0.005235	1
Common bean severe mosaic virus	Viruses; Geminiviridae	0.00519	2
Corchorus golden mosaic virus	Viruses; Geminiviridae	0.005352	2
Corchorus yellow spot virus	Viruses; Geminiviridae	0.005195	1
Corchorus yellow vein mosaic virus	Viruses; Geminiviridae	0.002743	1
Corchorus yellow vein virus – [HoaBinh]	Viruses; Geminiviridae	0.005415	1
Cotton chlorotic spot virus	Viruses; Geminiviridae	0.00532	1

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Cotton leaf crumple virus	Viruses; Geminiviridae	0.00518	2
Cotton leaf curl Alabad virus	Viruses; Geminiviridae	0.002744	4
Cotton leaf curl Allahabad virus [India: Karnal:OY77:2005]	Viruses; Geminiviridae	0.002744	1
Cotton leaf curl Bangalore virus	Viruses; Geminiviridae	0.002751	1
Cotton leaf curl Gezira virus	Viruses; Geminiviridae	0.00278	9
Cotton leaf curl Kokhran virus	Viruses; Geminiviridae	0.002759	3
Cotton leaf curl Multan virus	Viruses; Geminiviridae	0.002754	5
Cotton leaf curl Shahdadpur virus	Viruses; Geminiviridae	0.002748	1
Cotton leaf curl virus	Viruses; Geminiviridae	0.002753	1
Cotton yellow mosaic virus	Viruses; Geminiviridae	0.005482	2
Cowpea golden mosaic virus	Viruses; Geminiviridae	0.002728	1
Crassocephalum yellow vein virus – Jinghong	Viruses; Geminiviridae	0.002745	1
Cronartiumribicola	Eukaryota; Fungi; Basidiomycetes	94.3329	1
Croton yellow vein mosaic virus	Viruses; Geminiviridae	0.002757	2
Croton yellow vein virus	Viruses; Geminiviridae	0.002744	1
Cucurbit leaf crumple virus	Viruses; Geminiviridae	0.005232	1
'Cynodondactylon' phytoplasma	Bacteria; Terrabacteria group; Tenericutes	0.483935	1
Dalechampia chlorotic mosaic virus	Viruses; Geminiviridae	0.005214	1
Datura leaf curl virus	Viruses; Geminiviridae	0.002782	1
Datura leaf distortion virus	Viruses; Geminiviridae	0.005163	1
Deinbollia mosaic virus	Viruses; Geminiviridae	0.005461	1
Desmodium leaf distortion virus	Viruses; Geminiviridae	0.005083	1
Desmodium mottle virus	Viruses; Geminiviridae	0.00548	1
Dicliptera yellow mottle virus	Viruses; Geminiviridae	0.005204	2
Digitariaciliaris striate mosaic virus	Viruses; Geminiviridae	0.002816	2
Digitariadidactyla striate mosaic virus	Viruses; Geminiviridae	0.002762	1
Digitaria streak virus	Viruses; Geminiviridae	0.002701	1
Digitaria streak virus	Viruses; Geminiviridae	0.002701	1
Dolichos yellow mosaic virus	Viruses; Geminiviridae	0.005494	2
Dragonfly-associated mastrevirus	Viruses; Geminiviridae	0.00265	2
Duranta leaf curl virus	Viruses; Geminiviridae	0.002759	1
East African cassava mosaic Cameroon virus	Viruses; Geminiviridae	0.005543	1
East African cassava mosaic Kenya virus	Viruses; Geminiviridae	0.005573	1
East African cassava mosaic Malawi virus	Viruses; Geminiviridae	0.005558	2
East African cassava mosaic virus	Viruses; Geminiviridae	0.005576	4
East African cassava mosaic Zanzibar virus	Viruses; Geminiviridae	0.005548	1

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
'Echinacea purpurea' witches'-broom phytoplasma	Bacteria; Terrabacteria group; Tenericutes	0.545427	1
Eclipta yellow vein virus	Viruses; Geminiviridae	0.002748	2
Emilia sonchifolia yellow vein Thailand virus	Viruses; Geminiviridae	0.002746	1
Emilia yellow vein virus-[Fz1]	Viruses; Geminiviridae	0.002725	1
Eragrostiscurvula streak virus	Viruses; Geminiviridae	0.002754	2
Eragrostis minor streak virus	Viruses; Geminiviridae	0.002689	1
Eragrostis streak virus	Viruses; Geminiviridae	0.002746	1
Erectites yellow mosaic virus	Viruses; Geminiviridae	0.002751	1
Erwinia tracheiphila	Bacteria; Proteobacteria; Gammaproteobacteria	4.71727	3
Eupatorium yellow vein mosaic virus	Viruses; Geminiviridae	0.002778	1
Eupatorium yellow vein virus	Viruses; Geminiviridae	0.002767	5
Euphorbia caput-medusae latent virus	Viruses; Geminiviridae	0.002683	2
Euphorbia leaf curl Guangxi virus	Viruses; Geminiviridae	0.002747	1
Euphorbia leaf curl virus	Viruses; Geminiviridae	0.002746	1
Euphorbia mosaic Peru virus	Viruses; Geminiviridae	0.0026	1
Euphorbia mosaic virus	Viruses; Geminiviridae	0.005215	1
Euphorbia yellow leaf curl virus	Viruses; Geminiviridae	0.002731	1
Euphorbia yellow mosaic virus	Viruses; Geminiviridae	0.005187	2
Exomismicrophylla associated virus	Viruses; Geminiviridae	0.002974	1
French bean leaf curl virus	Viruses; Geminiviridae	0.002741	1
French bean severe leaf curl virus	Viruses; Geminiviridae	0.002771	1
Fusarium oxysporum	Eukaryota; Fungi; Ascomycetes	61.3869	129
Golovinomyces cichoracearum	Eukaryota; Fungi; Ascomycetes	65.8869	3
Gossypium darwinii symptomless virus	Viruses; Geminiviridae	0.00274	1
Gossypium punctatum mild leaf curl virus	Viruses; Geminiviridae	0.005462	1
Grapevine red blotch virus	Viruses; Geminiviridae	0.003206	2
Hedyotisuncinella yellow mosaic virus	Viruses; Geminiviridae	0.002749	1
Hemidesmus yellow mosaic virus	Viruses; Geminiviridae	0.002825	1
Hemileiavastatrix	Eukaryota; Fungi; Basidiomycetes	543.605	2
Hollyhock leaf crumple virus	Viruses; Geminiviridae	0.002755	1
Hollyhock leaf curl virus	Viruses; Geminiviridae	0.002748	1
Hollyhock yellow vein mosaic Islamabad virus	Viruses; Geminiviridae	0.002741	2
Hollyhock yellow vein mosaic virus	Viruses; Geminiviridae	0.00275	2
Honeysuckle yellow vein Kagoshima virus	Viruses; Geminiviridae	0.002762	1

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Honeysuckle yellow vein mosaic virus	Viruses; Geminiviridae	0.002759	2
Honeysuckle yellow vein virus	Viruses; Geminiviridae	0.002784	15
Horsegram yellow mosaic virus	Viruses; Geminiviridae	0.005405	1
Horseradish curly top virus	Viruses; Geminiviridae	0.00308	1
Indian cassava mosaic virus	Viruses; Geminiviridae	0.00546	4
Ipomoea yellow vein virus	Viruses; Geminiviridae	0.002791	1
Italian clover phyllody phytoplasma	Bacteria; Terrabacteria group; Tenericutes	0.597245	1
Jacquemontia mosaic Yucatan virus	Viruses; Geminiviridae	0.005193	1
Jacquemontia yellow mosaic virus	Viruses; Geminiviridae	0.005189	1
Jacquemontia yellow vein virus	Viruses; Geminiviridae	0.002585	1
Jatropha leaf crumple virus	Viruses; Geminiviridae	0.002735	1
Jatropha leaf curl Gujarat virus	Viruses; Geminiviridae	0.002758	1
Jatropha leaf curl virus	Viruses; Geminiviridae	0.002844	2
Jatropha leaf yellow mosaic Katarniaghat virus	Viruses; Geminiviridae	0.002744	1
Jatropha mosaic India virus	Viruses; Geminiviridae	0.00274	1
Jatropha mosaic Nigeria virus	Viruses; Geminiviridae	0.002781	1
Jatropha mosaic virus	Viruses; Geminiviridae	0.005198	1
Jatropha yellow mosaic virus	Viruses; Geminiviridae	0.002757	1
Kenaf leaf curl virus-[India: Bahraich:2007]	Viruses; Geminiviridae	0.00274	1
Kudzu mosaic virus	Viruses; Geminiviridae	0.005403	1
Leonurus mosaic virus	Viruses; Geminiviridae	0.002652	1
Linderniaanagallis yellow vein virus	Viruses; Geminiviridae	0.00274	1
Lisianthus enation leaf curl virus	Viruses; Geminiviridae	0.002759	1
Ludwigia yellow vein Vietnam virus	Viruses; Geminiviridae	0.002751	1
Ludwigia yellow vein virus	Viruses; Geminiviridae	0.002758	1
Luffa yellow mosaic virus	Viruses; Geminiviridae	0.005455	1
Lycianthes yellow mosaic virus	Viruses; Geminiviridae	0.005456	1
Macroptilium bright mosaic virus	Viruses; Geminiviridae	0.002636	1
Macroptilium common mosaic virus	Viruses; Geminiviridae	0.00523	1
Macroptilium golden mosaic virus	Viruses; Geminiviridae	0.005158	1
Macroptilium golden yellow mosaic virus	Viruses; Geminiviridae	0.005224	1
Macroptilium mosaic Puerto Rico virus	Viruses; Geminiviridae	0.005186	1
Macroptilium yellow mosaic Florida virus	Viruses; Geminiviridae	0.005247	1
Macroptilium yellow mosaic virus	Viruses; Geminiviridae	0.005223	2
Macroptilium yellow net virus	Viruses; Geminiviridae	0.005197	1
Macroptilium yellow spot virus	Viruses; Geminiviridae	0.00266	1
Macroptilium yellow vein virus	Viruses; Geminiviridae	0.002656	1
Maize bushy stunt phytoplasma	Bacteria; Terrabacteria group; Tenericutes	0.576118	1

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Maize streak Reunion virus	Viruses; Geminiviridae	0.002882	1
Maize streak Reunion virus	Viruses; Geminiviridae	0.002882	1
Maize streak virus	Viruses; Geminiviridae	0.002701	12
Maize striate mosaic virus	Viruses; Geminiviridae	0.002746	2
Malachra yellow mosaic virus	Viruses; Geminiviridae	0.002739	1
Malvastrum bright yellow mosaic virus	Viruses; Geminiviridae	0.005213	1
Malvastrum leaf curl Guangdong virus	Viruses; Geminiviridae	0.002767	1
Malvastrum leaf curl Philippines virus	Viruses; Geminiviridae	0.002742	1
Malvastrum leaf curl virus	Viruses; Geminiviridae	0.002745	1
Malvastrum yellow mosaic Helshire virus	Viruses; Geminiviridae	0.002609	1
Malvastrum yellow mosaic Jamaica virus	Viruses; Geminiviridae	0.005192	1
Malvastrum yellow mosaic virus	Viruses; Geminiviridae	0.002728	1
Malvastrum yellow vein Baoshan virus	Viruses; Geminiviridae	0.002745	2
Malvastrum yellow vein Cambodia virus	Viruses; Geminiviridae	0.002737	1
Malvastrum yellow vein Changa Manga virus	Viruses; Geminiviridae	0.002754	1
Malvastrum yellow vein Honghe virus	Viruses; Geminiviridae	0.00274	1
Malvastrum yellow vein virus	Viruses; Geminiviridae	0.002731	1
Malvastrum yellow vein Yunnan virus	Viruses; Geminiviridae	0.002747	1
Melochia mosaic virus	Viruses; Geminiviridae	0.005213	1
Melochia yellow mosaic virus	Viruses; Geminiviridae	0.005288	1
Meloidogyne arenaria	Eukaryota; Animals; Roundworms	284.032	3
Meloidogyne enterolobii	Eukaryota; Animals; Roundworms	162.967	1
Meloidogyne floridensis	Eukaryota; Animals; Roundworms	74.846	2
Meloidogyne graminicola	Eukaryota; Animals; Roundworms	38.185	1
Meloidogyne hapla	Eukaryota; Animals; Roundworms	53.013	1
Meloidogyne incognita	Eukaryota; Animals; Roundworms	183.532	3
Meloidogyne javanica	Eukaryota; Animals; Roundworms	150.345	2
Melon chlorotic leaf curl virus	Viruses; Geminiviridae	0.005325	3
Merremia mosaic Puerto Rico virus	Viruses; Geminiviridae	0.005225	1
Merremia mosaic virus	Viruses; Geminiviridae	0.005085	2
Mesta yellow vein mosaic Bahraich virus	Viruses; Geminiviridae	0.002737	1
Mesta yellow vein mosaic virus	Viruses; Geminiviridae	0.002752	2
Milkweed yellows phytoplasma	Bacteria; Terrabacteria group; Tenericutes	0.583806	1

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Mimosa yellow leaf curl virus	Viruses; Geminiviridae	0.002757	1
Mirabilis leaf curl virus	Viruses; Geminiviridae	0.002778	1
Miscanthus streak virus	Viruses; Geminiviridae	0.002672	1
Miscanthus streak virus	Viruses; Geminiviridae	0.002672	1
Mungbean yellow mosaic India virus	Viruses; Geminiviridae	0.005361	1
Mungbean yellow mosaic virus	Viruses; Geminiviridae	0.005398	1
New Jersey aster yellows phytoplasma	Bacteria; Terrabacteria group; Tenericutes	0.652092	1
Oat dwarf virus	Viruses; Geminiviridae	0.00274	1
Oat dwarf virus	Viruses; Geminiviridae	0.00274	1
Okra enation leaf curl virus	Viruses; Geminiviridae	0.002738	1
Okra enation leaf curl virus [India: Munthal EL37:2006]	Viruses; Geminiviridae	0.002724	1
Okra leaf curl Cameroon virus	Viruses; Geminiviridae	0.002764	1
Okra leaf curl India virus [India:Sonipat EL14A:2006]	Viruses; Geminiviridae	0.002723	1
Okra leaf curl Oman virus	Viruses; Geminiviridae	0.002788	1
Okra leaf curl virus	Viruses; Geminiviridae	0.002386	1
Okra mottle virus	Viruses; Geminiviridae	0.005313	1
Okra yellow crinkle virus	Viruses; Geminiviridae	0.002795	3
Okra yellow mosaic Mexico virus	Viruses; Geminiviridae	0.005194	1
Onion yellows phytoplasma	Bacteria; Terrabacteria group; Tenericutes	0.853092	1
Ophiognomoniaclavigignenti-juglandacearum	Eukaryota; Fungi; Ascomycetes	52.5149	3
Oxalis yellow vein virus	Viruses; Geminiviridae	0.002661	1
Panicum streak virus	Viruses; Geminiviridae	0.002736	9
Papaya leaf crumple virus-Panipat 8 [India:Panipat:Papaya:2008]	Viruses; Geminiviridae	0.002736	1
Papaya leaf curl China virus	Viruses; Geminiviridae	0.002751	5
Papaya leaf curl Guandong virus	Viruses; Geminiviridae	0.002764	2
Papaya leaf curl virus	Viruses; Geminiviridae	0.002769	12
Paspalum dilatatum striate mosaic virus	Viruses; Geminiviridae	0.002806	1
Paspalum striate mosaic virus	Viruses; Geminiviridae	0.002816	2
Passionfruit leaf distortion virus	Viruses; Geminiviridae	0.005172	1
Passionfruit severe leaf distortion virus	Viruses; Geminiviridae	0.005316	1
Pavonia mosaic virus	Viruses; Geminiviridae	0.005367	1
Pavonia yellow mosaic virus	Viruses; Geminiviridae	0.005378	1
Pea leaf distortion virus	Viruses; Geminiviridae	0.002738	1
Peanut witches'-broom phytoplasma	Bacteria; Terrabacteria group; Tenericutes	0.566694	1
Pectobacterium actinidiae	Bacteria; Proteobacteria; Gammaproteobacteria	4.92217	3

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Pectobacterium aquaticum	Bacteria; Proteobacteria; Gammaproteobacteria	4.46724	6
Pectobacterium atrosepticum	Bacteria; Proteobacteria; Gammaproteobacteria	5.10459	11
Pectobacterium betavasculorum	Bacteria; Proteobacteria; Gammaproteobacteria	4.68521	2
Pectobacterium brasiliense	Bacteria; Proteobacteria; Gammaproteobacteria	5.02631	28
Pectobacterium carotovorum	Bacteria; Proteobacteria; Gammaproteobacteria	4.86291	33
Pectobacterium fontis	Bacteria; Proteobacteria; Gammaproteobacteria	4.15156	1
Pectobacterium odoriferum	Bacteria; Proteobacteria; Gammaproteobacteria	5.4726	15
Pectobacterium parmentieri	Bacteria; Proteobacteria; Gammaproteobacteria	5.2273	19
Pectobacterium peruviense	Bacteria; Proteobacteria; Gammaproteobacteria	4.87102	5
Pectobacterium polaris	Bacteria; Proteobacteria; Gammaproteobacteria	5.00842	7
Pectobacterium polonicum	Bacteria; Proteobacteria; Gammaproteobacteria	4.83613	1
Pectobacterium punjabense	Bacteria; Proteobacteria; Gammaproteobacteria	4.73253	1
Pectobacterium versatile	Bacteria; Proteobacteria; Gammaproteobacteria	4.94937	1
Pectobacterium wasabiae	Bacteria; Proteobacteria; Gammaproteobacteria	5.1493	4
Pectobacterium zantedeschiae	Bacteria; Proteobacteria; Gammaproteobacteria	5.094	3
Pedilanthus leaf curl virus	Viruses; Geminiviridae	0.002764	3
Pepper golden mosaic virus	Viruses; Geminiviridae	0.005208	3

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Pepper huasteco yellow vein virus	Viruses; Geminiviridae	0.00522	1
Pepper leaf curl Bangladesh virus	Viruses; Geminiviridae	0.002754	3
Pepper leaf curl Lahore virus	Viruses; Geminiviridae	0.00274	1
Pepper leaf curl Lahore Virus-[Pakistan:Lahore1:2004]	Viruses; Geminiviridae	0.002747	1
Pepper leaf curl virus	Viruses; Geminiviridae	0.00276	3
Pepper leaf curl Yunnan virus-[YN323]	Viruses; Geminiviridae	0.002747	1
Pepper leafroll virus	Viruses; Geminiviridae	0.002568	1
Pepper yellow dwarf virus – Mexico	Viruses; Geminiviridae	0.002971	1
Pepper yellow dwarf virus – New Mexico	Viruses; Geminiviridae	0.002959	1
Pepper yellow leaf curl Indonesia virus	Viruses; Geminiviridae	0.005476	1
Pepper yellow leaf curl Thailand virus	Viruses; Geminiviridae	0.005474	2
Pepper yellow leaf curl virus	Viruses; Geminiviridae	0.006028	2
Pepper yellow leaf curl virus PSSWS-14	Viruses; Geminiviridae	0.002748	1
Pepper yellow vein Mali virus	Viruses; Geminiviridae	0.002786	1
Periwinkle leaf yellowing phytoplasma	Bacteria; Terrabacteria group; Tenericutes	0.824596	1
Phytophthora infestans	Eukaryota; Protists; Other Protists	228.544	2
Plantago lanceolata latent virus	Viruses; Geminiviridae	0.002832	1
Plasmodiophora brassicae	Eukaryota; Protists; Other Protists	24.5596	7
Poinsettia branch-inducing phytoplasma	Bacteria; Terrabacteria group; Tenericutes	0.63144	1
Potato yellow mosaic Panama virus	Viruses; Geminiviridae	0.005126	1
Potato yellow mosaic virus	Viruses; Geminiviridae	0.00514	4
Pouzolzia golden mosaic virus	Viruses; Geminiviridae	0.002725	2
Pouzolzia mosaic Guangdong virus	Viruses; Geminiviridae	0.002739	1
Premna leaf curl virus	Viruses; Geminiviridae	0.002753	1
Prunus latent virus	Viruses; Geminiviridae	0.003174	1
Pseudomonas syringae	Bacteria; Proteobacteria; Gammaproteobacteria	6.0937	376
Pseudoperonospora cubensis	Eukaryota; Protists; Other Protists	64.3328	1
Pumpkin yellow mosaic Malaysia virus	Viruses; Geminiviridae	0.002724	1
Pythium aphanidermatum	Eukaryota; Protists; Other Protists	35.8768	1
Pythium arrhenomanes	Eukaryota; Protists; Other Protists	44.6726	1
Pythium brassicum	Eukaryota; Protists; Other Protists	50.0694	1
Pythium guiyangense	Eukaryota; Protists; Other Protists	110.178	1

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Pythium insidiosum	Eukaryota; Protists; Other Protists	53.239	12
Pythium irregulare	Eukaryota; Protists; Other Protists	42.9681	2
Pythium iwayamai	Eukaryota; Protists; Other Protists	43.1992	1
Pythium oligandrum	Eukaryota; Protists; Other Protists	41.9689	3
Pythium periplocum	Eukaryota; Protists; Other Protists	35.8865	1
Pythium splendens	Eukaryota; Protists; Other Protists	53.361	1
Radish leaf curl virus	Viruses; Geminiviridae	0.002759	2
Ramie mosaic virus	Viruses; Geminiviridae	0.005446	1
Ramie mosaic Yunnan virus	Viruses; Geminiviridae	0.002759	1
Rhynchosia golden mosaic Havana virus-[Cuba:Havana:28:2007]	Viruses; Geminiviridae	0.005151	1
Rhynchosia golden mosaic Sinaloa virus	Viruses; Geminiviridae	0.005103	1
Rhynchosia golden mosaic virus	Viruses; Geminiviridae	0.005174	3
Rhynchosia golden mosaic Yucatan virus	Viruses; Geminiviridae	0.005139	1
Rhynchosia mild mosaic virus	Viruses; Geminiviridae	0.005162	1
Rhynchosia rugose golden mosaic virus	Viruses; Geminiviridae	0.005186	1
Rhynchosia yellow mosaic India virus	Viruses; Geminiviridae	0.005406	1
Rhynchosia yellow mosaic virus	Viruses; Geminiviridae	0.005379	1
Rice latent virus 1	Viruses; Geminiviridae	0.002757	2
Rice latent virus 2	Viruses; Geminiviridae	0.002843	1
Rice orange leaf phytoplasma	Bacteria; Terrabacteriagroup; Tenericutes	0.599264	1
Rose leaf curl virus	Viruses; Geminiviridae	0.002741	1
Saccharum streak virus	Viruses; Geminiviridae	0.002744	1
Sauropus leaf curl virus	Viruses; Geminiviridae	0.002762	1
Senecio yellow mosaic virus	Viruses; Geminiviridae	0.002746	1
Senna leaf curl virus	Viruses; Geminiviridae	0.002742	1
Sida angular mosaic virus	Viruses; Geminiviridae	0.005349	1
Sida bright yellow mosaic virus	Viruses; Geminiviridae	0.005348	1
Sida chlorotic mottle virus	Viruses; Geminiviridae	0.002601	1
Sida chlorotic vein virus	Viruses; Geminiviridae	0.005145	1
Sidaciliaris golden mosaic virus	Viruses; Geminiviridae	0.002638	1
Sida common mosaic virus	Viruses; Geminiviridae	0.002687	1
Sida golden mosaic Braco virus	Viruses; Geminiviridae	0.0026	1
Sida golden mosaic Brazil virus	Viruses; Geminiviridae	0.002659	1

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Sida golden mosaic Buckup virus-[Jamaica:St. Elizabeth:2004]	Viruses; Geminiviridae	0.005199	1
Sida golden mosaic Costa Rica virus	Viruses; Geminiviridae	0.005192	1
Sida golden mosaic Florida virus	Viruses; Geminiviridae	0.005186	2
Sida golden mosaic Honduras virus	Viruses; Geminiviridae	0.005192	1
Sida golden mosaic Lara virus	Viruses; Geminiviridae	0.002633	1
Sida golden mosaic virus	Viruses; Geminiviridae	0.005227	1
Sida golden mottle virus	Viruses; Geminiviridae	0.005184	1
Sida golden yellow spot virus	Viruses; Geminiviridae	0.002813	1
Sida golden yellow vein virus	Viruses; Geminiviridae	0.002603	1
Sida golden yellow vein virus-[Jamaica: Liguanea2:2008]	Viruses; Geminiviridae	0.00515	1
Sida leaf curl virus	Viruses; Geminiviridae	0.002757	1
Sida micrantha mosaic virus	Viruses; Geminiviridae	0.005331	4
Sida mosaic Alagoas virus	Viruses; Geminiviridae	0.005292	1
Sida mosaic Bolivia virus 1	Viruses; Geminiviridae	0.005348	1
Sida mosaic Bolivia virus 2	Viruses; Geminiviridae	0.005316	1
Sida mosaic Sinaloa virus	Viruses; Geminiviridae	0.005182	1
Sida mottle Alagoas virus	Viruses; Geminiviridae	0.002649	1
Sida mottle virus	Viruses; Geminiviridae	0.002668	1
Sida yellow blotch virus	Viruses; Geminiviridae	0.002664	1
Sida yellow leaf curl virus	Viruses; Geminiviridae	0.002664	1
Sida yellow mosaic Alagoas virus	Viruses; Geminiviridae	0.00269	1
Sida yellow mosaic China virus	Viruses; Geminiviridae	0.002751	1
Sida yellow mosaic virus	Viruses; Geminiviridae	0.002661	1
Sida yellow mosaic Yucatan virus	Viruses; Geminiviridae	0.005197	1
Sida yellow mottle virus	Viruses; Geminiviridae	0.005222	1
Sida yellow net virus	Viruses; Geminiviridae	0.002676	1
Sida yellow vein Madurai virus	Viruses; Geminiviridae	0.002753	1
Sida yellow vein virus	Viruses; Geminiviridae	0.005205	1
Sida strum golden leaf spot virus	Viruses; Geminiviridae	0.002666	1
Siegesbeckia yellow vein Guangxi virus	Viruses; Geminiviridae	0.002784	1
Siegesbeckia yellow vein virus	Viruses; Geminiviridae	0.002768	1
Solanum mosaic Bolivia virus	Viruses; Geminiviridae	0.005196	1
South African cassava mosaic virus	Viruses; Geminiviridae	0.00556	1
Soybean blistering mosaic virus	Viruses; Geminiviridae	0.002605	1
Soybean chlorotic blotch virus	Viruses; Geminiviridae	0.005355	1
Soybean chlorotic spot virus	Viruses; Geminiviridae	0.005208	1
Soybean mild mottle virus	Viruses; Geminiviridae	0.002768	1
Spilanthes yellow vein virus	Viruses; Geminiviridae	0.002761	1
Spinach curly top Arizona virus	Viruses; Geminiviridae	0.00286	1
Spinach severe curly top virus	Viruses; Geminiviridae	0.003065	1
Spinach yellow vein Sikar virus	Viruses; Geminiviridae	0.002753	1

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Sporisorium scitamineum	Eukaryota; Fungi; Basidiomycetes	20.0676	4
Sporobolus striate mosaic virus 1	Viruses; Geminiviridae	0.002789	1
Sporobolus striate mosaic virus 1	Viruses; Geminiviridae	0.002789	1
Sporobolus striate mosaic virus 2	Viruses; Geminiviridae	0.002716	1
Squash leaf curl China virus	Viruses; Geminiviridae	0.002756	3
Squash leaf curl China virus – [B]	Viruses; Geminiviridae	0.005455	1
Squash leaf curl Philippines virus	Viruses; Geminiviridae	0.005444	1
Squash leaf curl virus	Viruses; Geminiviridae	0.005241	1
Squash leaf curl Yunnan virus	Viruses; Geminiviridae	0.002714	1
Squash mild leaf curl virus	Viruses; Geminiviridae	0.00519	1
Sri Lankan cassava mosaic virus	Viruses; Geminiviridae	0.005466	2
Stachytarpheta leaf curl virus	Viruses; Geminiviridae	0.002749	1
Streptomyces scabiei	Bacteria; Terrabacteria group; Actinobacteria	10.1487	17
Sugarcane chlorotic streak virus	Viruses; Geminiviridae	0.002757	1
Sugarcane streak Egypt virus	Viruses; Geminiviridae	0.002706	1
Sugarcane streak Reunion virus	Viruses; Geminiviridae	0.00274	2
Sugarcane streak virus	Viruses; Geminiviridae	0.002758	2
Sugarcane striate virus	Viruses; Geminiviridae	0.002749	2
Sugarcane white streak virus	Viruses; Geminiviridae	0.00283	1
Sunn hemp leaf distortion virus	Viruses; Geminiviridae	0.002774	1
Sweet potato golden vein associated virus	Viruses; Geminiviridae	0.002824	1
Sweet potato golden vein Korea virus	Viruses; Geminiviridae	0.002807	1
Sweet potato leaf curl Bengal virus	Viruses; Geminiviridae	0.002823	1
Sweet potato leaf curl Canary virus	Viruses; Geminiviridae	0.002837	2
Sweet potato leaf curl China virus	Viruses; Geminiviridae	0.002771	1
Sweet potato leaf curl Georgia virus	Viruses; Geminiviridae	0.002773	1
Sweet potato leaf curl Guangxi virus	Viruses; Geminiviridae	0.002831	1
Sweet potato leaf curl Henan virus	Viruses; Geminiviridae	0.002785	2
Sweet potato leaf curl Lanzarote virus	Viruses; Geminiviridae	0.002814	1
Sweet potato leaf curl Sao Paulo virus	Viruses; Geminiviridae	0.002782	1
Sweet potato leaf curl Shanghai virus	Viruses; Geminiviridae	0.002834	1
Sweet potato leaf curl Sichuan virus 1	Viruses; Geminiviridae	0.002764	1
Sweet potato leaf curl Sichuan virus 2	Viruses; Geminiviridae	0.002786	1
Sweet potato leaf curl South Carolina virus	Viruses; Geminiviridae	0.002782	1
Sweet potato leaf curl Spain virus	Viruses; Geminiviridae	0.00278	1
Sweet potato leaf curl Uganda virus-[Uganda:Kampala:2008]	Viruses; Geminiviridae	0.002799	1
Sweet potato leaf curl virus	Viruses; Geminiviridae	0.002844	15
Sweet potato mosaic virus	Viruses; Geminiviridae	0.002803	2
Sweet potato symptomless virus 1	Viruses; Geminiviridae	0.002886	2

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Sweet potato symptomless virus 1	Viruses; Geminiviridae	0.002886	2
Switchgrass mosaic-associated virus 1	Viruses; Geminiviridae	0.002739	1
Synedrella leaf curl virus	Viruses; Geminiviridae	0.002749	1
Synedrella yellow vein clearing virus	Viruses; Geminiviridae	0.002751	1
Telfairia golden mosaic virus	Viruses; Geminiviridae	0.002742	1
Tilletia caries	Eukaryota; Fungi; Basidiomycetes	29.5409	2
Tobacco curly shoot virus	Viruses; Geminiviridae	0.002743	1
Tobacco leaf curl Comoros virus	Viruses; Geminiviridae	0.002755	1
Tobacco leaf curl Cuba virus	Viruses; Geminiviridae	0.005176	2
Tobacco leaf curl Japan virus	Viruses; Geminiviridae	0.002761	1
Tobacco leaf curl Pusa virus	Viruses; Geminiviridae	0.002707	1
Tobacco leaf curl Thailand virus	Viruses; Geminiviridae	0.002752	1
Tobacco leaf curl virus	Viruses; Geminiviridae	0.002762	1
Tobacco leaf curl Yunnan virus	Viruses; Geminiviridae	0.00275	1
Tobacco leaf curl Zimbabwe virus	Viruses; Geminiviridae	0.002767	1
Tobacco leaf rugose virus	Viruses; Geminiviridae	0.002622	1
Tobacco mosaic virus	Viruses; Virgaviridae	0.006395	1
Tobacco mottle leaf curl virus	Viruses; Geminiviridae	0.002634	1
Tobacco yellow crinkle virus	Viruses; Geminiviridae	0.005154	1
Tobacco yellow dwarf virus	Viruses; Geminiviridae	0.00258	1
Tobacco yellow dwarf virus	Viruses; Geminiviridae	0.00258	1
Tomato bright yellow mosaic virus	Viruses; Geminiviridae	0.002619	1
Tomato bright yellow mottle virus	Viruses; Geminiviridae	0.002639	1
Tomato chino La Paz virus	Viruses; Geminiviridae	0.002632	3
Tomato chlorotic leaf distortion virus-[Venezuela:Zulia:2004]	Viruses; Geminiviridae	0.00523	1
Tomato chlorotic mottle Guyane virus	Viruses; Geminiviridae	0.005234	1
Tomato chlorotic mottle virus	Viruses; Geminiviridae	0.005195	3
Tomato common mosaic virus	Viruses; Geminiviridae	0.005058	1
Tomato curly stunt virus	Viruses; Geminiviridae	0.002766	1
Tomato dwarf leaf virus	Viruses; Geminiviridae	0.005034	1
Tomato enation leaf curl virus	Viruses; Geminiviridae	0.002756	1
Tomato golden leaf distortion virus	Viruses; Geminiviridae	0.00263	1
Tomato golden leaf spot virus	Viruses; Geminiviridae	0.002669	1
Tomato golden mosaic virus	Viruses; Geminiviridae	0.005096	1
Tomato golden mottle virus	Viruses; Geminiviridae	0.005172	1
Tomato golden vein virus	Viruses; Geminiviridae	0.005095	1
Tomato interveinal chlorosis virus	Viruses; Geminiviridae	0.002617	1
Tomato latent virus	Viruses; Geminiviridae	0.002746	1
Tomato leaf curl Anjouan virus	Viruses; Geminiviridae	0.002781	1
Tomato leaf curl Arusha virus	Viruses; Geminiviridae	0.002766	2
Tomato leaf curl Bangalore virus	Viruses; Geminiviridae	0.002759	5

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Tomato leaf curl Bangladesh virus	Viruses; Geminiviridae	0.002761	1
Tomato leaf curl Barka virus	Viruses; Geminiviridae	0.002753	1
Tomato leaf curl Burkina Faso virus	Viruses; Geminiviridae	0.002784	1
Tomato leaf curl Cameroon virus	Viruses; Geminiviridae	0.002808	1
Tomato leaf curl Cebu virus	Viruses; Geminiviridae	0.002723	1
Tomato leaf curl China virus	Viruses; Geminiviridae	0.002738	4
Tomato leaf curl China virus – OX2	Viruses; Geminiviridae	0.002744	1
Tomato leaf curl Comoros virus	Viruses; Geminiviridae	0.002765	1
Tomato leaf curl Cotabato virus	Viruses; Geminiviridae	0.00275	1
Tomato leaf curl Diana virus	Viruses; Geminiviridae	0.002745	1
Tomato leaf curl Gandhinagar virus	Viruses; Geminiviridae	0.00276	1
Tomato leaf curl Ghana virus	Viruses; Geminiviridae	0.002803	2
Tomato leaf curl Guangdong virus	Viruses; Geminiviridae	0.002744	1
Tomato leaf curl Guangxi virus	Viruses; Geminiviridae	0.002752	1
Tomato leaf curl Gujarat virus	Viruses; Geminiviridae	0.005445	1
Tomato leaf curl Hainan virus	Viruses; Geminiviridae	0.002756	3
Tomato leaf curl Hanoi virus	Viruses; Geminiviridae	0.00274	1
Tomato leaf curl Iran virus	Viruses; Geminiviridae	0.002763	1
Tomato leaf curl Java virus	Viruses; Geminiviridae	0.002752	2
Tomato leaf curl Joydebpur virus	Viruses; Geminiviridae	0.002798	2
Tomato leaf curl Karnataka virus	Viruses; Geminiviridae	0.002772	7
Tomato leaf curl Kerala virus	Viruses; Geminiviridae	0.002767	2
Tomato leaf curl Kumasi virus	Viruses; Geminiviridae	0.002794	1
Tomato leaf curl Laos virus	Viruses; Geminiviridae	0.002748	1
Tomato leaf curl Liwa virus	Viruses; Geminiviridae	0.002761	1
Tomato leaf curl Madagascar virus	Viruses; Geminiviridae	0.002775	1
Tomato leaf curl Madagascar virus-Menabe [Madagascar: Morondova:2001]	Viruses; Geminiviridae	0.002777	1
Tomato leaf curl Malaysia virus	Viruses; Geminiviridae	0.002754	1
Tomato leaf curl Mali virus	Viruses; Geminiviridae	0.002773	1
Tomato leaf curl Mayotte virus	Viruses; Geminiviridae	0.002768	1
Tomato leaf curl Mindanao virus	Viruses; Geminiviridae	0.002761	1
Tomato leaf curl Moheli virus	Viruses; Geminiviridae	0.002756	1
Tomato leaf curl Namakely virus	Viruses; Geminiviridae	0.002772	2
Tomato leaf curl New Delhi virus	Viruses; Geminiviridae	0.005435	8
Tomato leaf curl New Delhi virus 2	Viruses; Geminiviridae	0.002735	1
Tomato leaf curl New Delhi virus 4	Viruses; Geminiviridae	0.002739	1
Tomato leaf curl Nigeria virus-[Nigeria:2006]	Viruses; Geminiviridae	0.002784	1
Tomato leaf curl Oman virus	Viruses; Geminiviridae	0.002763	1
Tomato leaf curl Palampur virus	Viruses; Geminiviridae	0.005481	4

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Tomato leaf curl Patna virus	Viruses; Geminiviridae	0.002752	1
Tomato leaf curl Philippines virus	Viruses; Geminiviridae	0.002755	3
Tomato leaf curl Pune virus	Viruses; Geminiviridae	0.002756	1
Tomato leaf curl purple vein virus	Viruses; Geminiviridae	0.002629	1
Tomato leaf curl Rajasthan virus	Viruses; Geminiviridae	0.002758	1
Tomato leaf curl Ranchi virus	Viruses; Geminiviridae	0.002762	1
Tomato leaf curl Seychelles virus	Viruses; Geminiviridae	0.002742	1
Tomato leaf curl Sinaloa virus	Viruses; Geminiviridae	0.005173	1
Tomato leaf curl Sri Lanka virus	Viruses; Geminiviridae	0.002756	1
Tomato leaf curl Sudan virus	Viruses; Geminiviridae	0.002782	4
Tomato leaf curl Sulawesi virus	Viruses; Geminiviridae	0.002751	1
Tomato leaf curl Taiwan virus	Viruses; Geminiviridae	0.002743	4
Tomato leaf curl Toliara virus	Viruses; Geminiviridae	0.002764	1
Tomato leaf curl Uganda virus	Viruses; Geminiviridae	0.002747	1
Tomato leaf curl Vietnam virus	Viruses; Geminiviridae	0.002745	1
Tomato leaf curl virus	Viruses; Geminiviridae	0.002766	5
Tomato leaf deformation virus	Viruses; Geminiviridae	0.002591	1
Tomato leaf distortion virus	Viruses; Geminiviridae	0.002645	1
Tomato mild mosaic virus	Viruses; Geminiviridae	0.005371	1
Tomato mild yellow leaf curl Aragua virus	Viruses; Geminiviridae	0.005168	1
Tomato mosaic Havana virus	Viruses; Geminiviridae	0.005206	1
Tomato mosaic Trujillo virus	Viruses; Geminiviridae	0.002637	1
Tomato mottle leaf curl virus	Viruses; Geminiviridae	0.005229	2
Tomato mottle Taino virus	Viruses; Geminiviridae	0.005159	1
Tomato mottle virus	Viruses; Geminiviridae	0.005145	1
Tomato mottle wrinkle virus	Viruses; Geminiviridae	0.005124	1
Tomato pseudo-curly top virus	Viruses; Geminiviridae	0.002861	1
Tomato rugose mosaic virus	Viruses; Geminiviridae	0.005194	1
Tomato rugose yellow leaf curl virus	Viruses; Geminiviridae	0.005305	1
Tomato severe leaf curl virus	Viruses; Geminiviridae	0.002755	4
Tomato severe rugose virus	Viruses; Geminiviridae	0.005164	1
Tomato spotted wilt tospovirus	Viruses; Tospoviridae	0.016634	1
Tomato yellow leaf curl Axarquia virus	Viruses; Geminiviridae	0.002763	1
Tomato yellow leaf curl China virus	Viruses; Geminiviridae	0.002741	7
Tomato yellow leaf curl Guangdong virus	Viruses; Geminiviridae	0.002744	1
Tomato yellow leaf curl Indonesia virus-[Lembang]	Viruses; Geminiviridae	0.002762	1
Tomato yellow leaf curl Kanchanaburi virus	Viruses; Geminiviridae	0.005504	1
Tomato yellow leaf curl Malaga virus	Viruses; Geminiviridae	0.002782	1

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Tomato yellow leaf curl Mali virus	Viruses; Geminiviridae	0.002796	3
Tomato yellow leaf curl Sardinia virus	Viruses; Geminiviridae	0.002773	1
Tomato yellow leaf curl Saudi virus	Viruses; Geminiviridae	0.002775	1
Tomato yellow leaf curl Shuangbai virus – [Y4536]	Viruses; Geminiviridae	0.002748	1
Tomato yellow leaf curl Thailand virus	Viruses; Geminiviridae	0.005488	5
Tomato yellow leaf curl Vietnam virus	Viruses; Geminiviridae	0.002745	1
Tomato yellow leaf curl virus	Viruses; Geminiviridae	0.00279	6
Tomato yellow leaf curl Yunnan virus	Viruses; Geminiviridae	0.002754	1
Tomato yellow leaf distortion virus	Viruses; Geminiviridae	0.005219	1
Tomato yellow margin leaf curl virus	Viruses; Geminiviridae	0.005118	1
Tomato yellow mottle virus	Viruses; Geminiviridae	0.005121	1
Tomato yellow spot virus	Viruses; Geminiviridae	0.0053	1
Tomato yellow vein streak virus	Viruses; Geminiviridae	0.00513	1
Triumfetta yellow mosaic virus	Viruses; Geminiviridae	0.005277	1
Turnip curly top virus	Viruses; Geminiviridae	0.002981	4
Turnip curly top virus	Viruses; Geminiviridae	0.002981	4
Turnip leaf roll virus	Viruses; Geminiviridae	0.002965	1
Turnip leaf roll virus	Viruses; Geminiviridae	0.002965	1
TYLCAxV-Sic1-[IT:Sic2/2:04]	Viruses; Geminiviridae	0.002771	1
Urochloa streak virus	Viruses; Geminiviridae	0.002736	1
Urochloa streak virus	Viruses; Geminiviridae	0.002736	1
Ustilago maydis	Eukaryota; Fungi; Basidiomycetes	19.6644	7
Vaccinium witches'-broom phytoplasma	Bacteria; Terrabacteriagroup; Tenericutes	0.647754	1
Velvet bean golden mosaic virus	Viruses; Geminiviridae	0.002767	1
Velvet bean severe mosaic virus	Viruses; Geminiviridae	0.00539	1
Venturia inaequalis	Eukaryota; Fungi; Ascomycetes	72.7916	85
Vernonia crinkle virus	Viruses; Geminiviridae	0.002791	1
Vernonia yellow vein Fujian virus	Viruses; Geminiviridae	0.002739	1
Vernonia yellow vein virus	Viruses; Geminiviridae	0.002745	1
Verticillium albo-atrum	Eukaryota; Fungi; Ascomycetes	36.4685	1
Verticillium alfalfae	Eukaryota; Fungi; Ascomycetes	32.863	2
Verticillium dahliae	Eukaryota; Fungi; Ascomycetes	33.9003	13
Verticillium isaacii	Eukaryota; Fungi; Ascomycetes	35.6909	1
Verticillium klebahnii	Eukaryota; Fungi; Ascomycetes	36.0824	1

(continued)

Table 32.2 (continued)

Organism name	Organism groups	Size(Mb)	Assemblies
Verticillium longisporum	Eukaryota; Fungi; Ascomycetes	99.1892	2
Verticillium nonalfalfae	Eukaryota; Fungi; Ascomycetes	31.7515	3
Verticillium nubilum	Eukaryota; Fungi; Ascomycetes	37.9116	1
Verticillium tricorpus	Eukaryota; Fungi; Ascomycetes	36.0604	2
Verticillium zaregamsianum	Eukaryota; Fungi; Ascomycetes	37.1319	1
Vigna yellow mosaic virus	Viruses; Geminiviridae	0.002602	1
Vinca leaf curl virus	Viruses; Geminiviridae	0.002776	1
Watermelon chlorotic stunt virus	Viruses; Geminiviridae	0.005498	1
West African Asystasia virus 1	Viruses; Geminiviridae	0.005388	2
West African Asystasia virus 2	Viruses; Geminiviridae	0.002744	1
Wheat blue dwarf phytoplasma	Bacteria; Terrabacteriagroup; Tenericutes	0.611462	1
Wheat dwarf India virus	Viruses; Geminiviridae	0.002783	1
Wheat dwarf virus	Viruses; Geminiviridae	0.00275	6
Wheat dwarf virus	Viruses; Geminiviridae	0.00275	6
Whitefly-associated begomovirus 1	Viruses; Geminiviridae	0.002609	1
Whitefly-associated begomovirus 2	Viruses; Geminiviridae	0.00259	1
Whitefly-associated begomovirus 3	Viruses; Geminiviridae	0.002629	1
Whitefly-associated begomovirus 4	Viruses; Geminiviridae	0.002608	1
Whitefly-associated begomovirus 6	Viruses; Geminiviridae	0.002638	1
Whitefly-associated begomovirus 7	Viruses; Geminiviridae	0.002767	1
Wissadula golden mosaic virus	Viruses; Geminiviridae	0.0052	1
Wissadula yellow mosaic virus	Viruses; Geminiviridae	0.002621	1
Xanthomonas axonopodis	Bacteria; Proteobacteria; Gammaproteobacteria	5.41058	14
Xanthomonas campestris	Bacteria; Proteobacteria; Gammaproteobacteria	5.07619	71

Table 32.3 Plant genome sequence details

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Abrus precatorius	Eukaryota; Plants; Land Plants	GCA_003935025.1	Scaffold	347.23	QYUI01	160	40048
Acer yangbiense	Eukaryota; Plants; Land Plants	GCA_008009225.1	Chromosome	665.888	VAHF01	280	28320
Actinidia chinensis	Eukaryota; Plants; Land Plants	GCA_000467755.1	Contig	604.217	AONS01	26721	0
Actinidia chinensis var. chinensis	Eukaryota; Plants; Land Plants	GCA_003024255.1	Chromosome	553.842	NKQK01	1234	33115
Actinidia eriantha	Eukaryota; Plants; Land Plants	GCA_004150315.1	Chromosome	690.611	QOVS01	1735	0
Aegilops tauschii	Eukaryota; Plants; Land Plants	GCA_000347335.2	Chromosome	4310.35	AOCO02	112210	0
Aegilops tauschii	Eukaryota; Plants; Land Plants	GCA_002105435.1	Chromosome	247.197	LYXL01	1	0
Aegilops tauschii subsp. strangulata	Eukaryota; Plants; Land Plants	GCA_002575655.1	Chromosome	4224.92	NWVB01	109583	0
Aegilops tauschii subsp. tauschii	Eukaryota; Plants; Land Plants	GCA_001957025.1	Contig	4327.32	MCGU01	68538	55713
Aethionema arabicum	Eukaryota; Plants; Land Plants	GCA_000411095.1	Scaffold	192.488	ASZG01	18312	0
Alloteropsis semialata	Eukaryota; Plants; Land Plants	GCA_004135705.1	Chromosome	747.772	QPGU01	688	0
Alnus glutinosa	Eukaryota; Plants; Land Plants	GCA_003254965.1	Scaffold	611.874	QAOD01	167345	0
Amaranthus hypochondriacus	Eukaryota; Plants; Land Plants	GCA_000753965.1	Scaffold	502.148	JPXE01	117340	0
Amaranthus tuberculatus	Eukaryota; Plants; Land Plants	GCA_000180655.1	Contig	4.34798	ACQK01	15440	0

Amborella trichopoda	Eukaryota; Plants; Land Plants	GCA_000471905.1	Scaffold	706.495	AWHE01	5746	31494
Ananas comosus	Eukaryota; Plants; Land Plants	GCA_902162155.1	Scaffold	315.839	CABGUK01	25	0
Ananas comosus	Eukaryota; Plants; Land Plants	GCA_001661175.1	Scaffold	524.07	LSRQ01	8448	23598
Ananas comosus	Eukaryota; Plants; Land Plants	GCA_001540865.1	Chromosome	382.056	LODP01	3129	35775
Ananas comosus var. bracteatus	Eukaryota; Plants; Land Plants	GCA_902506285.1	Scaffold	513.235	CABWKS01	103	0
Anastatica hierochuntica	Eukaryota; Plants; Land Plants	GCA_900406275.1	Scaffold	542.343	OVAN01	72649	0
Andrographis paniculata	Eukaryota; Plants; Land Plants	GCA_004354405.1	Chromosome	269.408	SMLO01	257	0
Apostasia shenzhenica	Eukaryota; Plants; Land Plants	GCA_002786265.1	Scaffold	348.733	PEFY01	2985	21743
Aquilaria agallochum	Eukaryota; Plants; Land Plants	GCA_000696445.1	Scaffold	726.71	JMHV01	27769	0
Aquilaria sinensis	Eukaryota; Plants; Land Plants	GCA_005392925.1	Contig	699.794	SMDT01	3368	0
Aquilegia coerulea	Eukaryota; Plants; Land Plants	GCA_002738505.1	Scaffold	301.98	NXFA01	970	41063
Arabidopsis halleri	Eukaryota; Plants; Land Plants	GCA_003711535.1	Scaffold	164.574	RCNM01	40344	0
Arabidopsis halleri subsp. gemmifera	Eukaryota; Plants; Land Plants	GCA_900078215.1	Scaffold	196.243	FJVB01	2239	0
Arabidopsis halleri subsp. gemmifera	Eukaryota; Plants; Land Plants	GCA_000523005.1	Scaffold	221.14	BASO01	282453	0
Arabidopsis halleri subsp. gemmifera	Eukaryota; Plants; Land Plants	GCA_003118655.1	Scaffold	413.881	BFAE01	344622	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Arabidopsis lyrata subsp. lyrata	Eukaryota; Plants; Land Plants	GCA_000004255.1	Scaffold	206.823	ADBK01	696	39161
Arabidopsis lyrata subsp. petraea	Eukaryota; Plants; Land Plants	GCA_900205625.1	Scaffold	175.183	OANL01	1675	0
Arabidopsis lyrata subsp. petraea	Eukaryota; Plants; Land Plants	GCA_000524985.1	Scaffold	202.972	BASP01	281536	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900303355.1	Contig	119.503	OMOL01	62	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_000835945.1	Contig	127.419	JSAD01	378	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900303345.1	Contig	119.75	OMOK01	78	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900243945.1	Contig	119.167	OFAM01	59	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900243935.1	Contig	119.203	OFEF01	40	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_001753755.2	Contig	244.583	MJMM01	411	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900243955.1	Contig	119.128	OFAN01	139	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_001742845.1	Scaffold	116.846	LXSY01	5197	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_000222345.1	Scaffold	98.0662	AFNB01	1740	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_000222325.1	Scaffold	96.5002	AFNA01	2143	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_000222365.1	Scaffold	96.2565	AFMZ01	1261	0

Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_000222385.1	Scaffold	96.694	AFNC01	2408	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234075.1	Contig	0.195283	OCZF01	19	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234235.1	Contig	1.71971	OCYH01	133	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233735.1	Contig	1.87192	OCWZ01	156	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234125.1	Contig	2.02645	OCXO01	164	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234305.1	Contig	2.01361	OCYE01	163	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233785.1	Contig	2.68429	OCWV01	231	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234265.1	Contig	1.60963	OCYD01	141	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234185.1	Contig	2.04223	OCYB01	169	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233775.1	Contig	1.84553	OCWC01	154	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234295.1	Contig	1.88946	OCYF01	171	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234275.1	Contig	1.95856	OCYJ01	167	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233705.1	Contig	1.53762	OCVZ01	126	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234025.1	Contig	1.4141	OCXK01	116	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233685.1	Contig	1.72997	OCWA01	149	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233725.1	Contig	2.27092	OCWQ01	190	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234165.1	Contig	2.13757	OCYN01	189	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234135.1	Contig	1.73315	OCXS01	143	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233965.1	Contig	1.78161	OCXD01	170	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234175.1	Contig	1.93182	OCYM01	173	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234115.1	Contig	1.8648	OCXN01	164	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233995.1	Contig	2.27278	OCXQ01	199	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234145.1	Contig	1.83575	OCXW01	156	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234285.1	Contig	2.15462	OCYA01	194	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234105.1	Contig	1.76999	OCXU01	156	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233755.1	Contig	1.9492	OCWR01	183	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233985.1	Contig	1.75614	OCXL01	156	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234215.1	Contig	1.86784	OCYC01	173	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234205.1	Contig	1.83401	OCYL01	182	0

Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234065.1	Contig	1.56864	OCXH01	151	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233765.1	Contig	1.97277	OCWY01	169	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233795.1	Contig	1.65792	OCXE01	153	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233745.1	Contig	1.86401	OCWJ01	172	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234155.1	Contig	1.93139	OCXX01	180	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234045.1	Contig	1.62901	OCXG01	158	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234195.1	Contig	1.93135	OCYG01	176	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233665.1	Contig	1.82675	OCWE01	191	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233875.1	Contig	1.72986	OCWP01	182	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233905.1	Contig	1.37218	OCWL01	132	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233635.1	Contig	1.68821	OCVY01	157	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233835.1	Contig	1.65938	OCWB01	159	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234085.1	Contig	2.05357	OCXM01	194	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234325.1	Contig	1.55025	OCYO01	153	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233645.1	Contig	1.34032	OCWG01	133	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234245.1	Contig	1.90115	OCXY01	200	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234095.1	Contig	1.7997	OCXR01	173	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233865.1	Contig	1.33758	OCWN01	133	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233805.1	Contig	1.5272	OCWS01	146	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234035.1	Contig	2.18075	OCXI01	205	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234365.1	Contig	1.60312	OCYP01	153	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233825.1	Contig	1.4618	OCWH01	147	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233855.1	Contig	1.61604	OCXC01	153	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233695.1	Contig	1.80119	OCXF01	169	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234315.1	Contig	1.66717	OCYK01	179	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233915.1	Contig	1.36257	OCXB01	136	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234055.1	Contig	1.76317	OCXV01	182	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233885.1	Contig	1.38144	OCWT01	140	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233655.1	Contig	1.15189	OCWF01	124	0

Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234005.1	Contig	1.33529	OCXJ01	137	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233815.1	Contig	1.71934	OCWO01	188	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234015.1	Contig	1.11018	OCXT01	124	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233675.1	Contig	1.03397	OCWK01	120	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233845.1	Contig	1.15	OCWM01	133	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233895.1	Contig	1.31957	OCWX01	158	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233715.1	Contig	1.14084	OCXA01	138	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233945.1	Contig	1.26394	OCWD01	151	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233925.1	Contig	2.03504	OCWI01	248	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233975.1	Contig	0.844674	OCXP01	119	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900233955.1	Contig	0.838703	OCWU01	109	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234225.1	Contig	0.32515	OCYI01	52	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900234255.1	Contig	0.847842	OCXZ01	146	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_902460285.1	Chromosome	120.338	CABPTM01	105	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_900660825.1	Chromosome	119.627	CAACVU01	109	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_902460305.1	Chromosome	122.202	CABPTJ01	184	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_902460275.1	Chromosome	119.75	CABPTK01	102	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_902460295.1	Chromosome	120.29	CABPTI01	94	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_902460315.1	Chromosome	120.795	CABPTL01	142	0
Arabidopsis thaliana	Eukaryota; Plants; Land Plants	GCA_001651475.1	Chromosome	118.891	LUHQ01	30	30837
Arabis alpina	Eukaryota; Plants; Land Plants	GCA_000612745.1	Contig	171.788	CBTM01	37680	0
Arabis alpina	Eukaryota; Plants; Land Plants	GCA_000733195.1	Chromosome	308.033	JNGA01	27779	23286
Arabis montbretiana	Eukaryota; Plants; Land Plants	GCA_001484125.1	Contig	199.12	LNCH01	28775	0
Arabis nordmanniana	Eukaryota; Plants; Land Plants	GCA_001484925.1	Scaffold	342.307	LNCG01	267228	0
Arachis duranensis	Eukaryota; Plants; Land Plants	GCA_001687015.1	Scaffold	1075.96	MAMN01	20214	0
Arachis duranensis	Eukaryota; Plants; Land Plants	GCA_000817695.2	Chromosome	1084.26	JQIN01	3189	52826
Arachis hypogaea	Eukaryota; Plants; Land Plants	GCA_003086295.2	Chromosome	2557.07	PIVG01	385	100775
Arachis hypogaea	Eukaryota; Plants; Land Plants	GCA_004170445.1	Chromosome	2551.68	SDMP01	29	101330
Arachis ipaensis	Eukaryota; Plants; Land Plants	GCA_000816755.2	Chromosome	1353.5	JQIO01	997	57621

Arachis monticola	Eukaryota; Plants; Land Plants	GCA_003063285.2	Chromosome	2618.65	QBTX01	6909	0
Argania spinosa	Eukaryota; Plants; Land Plants	GCA_003260245.1	Scaffold	670.097	QLOD01	75327	0
Aristotelia chilensis	Eukaryota; Plants; Land Plants	GCA_008921755.1	Scaffold	96.3544	VEXP01	42602	0
Aristotelia chilensis	Eukaryota; Plants; Land Plants	GCA_008126665.1	Scaffold	0.312713	VDCA01	93	0
Artemisia annua	Eukaryota; Plants; Land Plants	GCA_003112345.1	Scaffold	1792.86	PKPP01	39400	213
Artocarpus camansi	Eukaryota; Plants; Land Plants	GCA_002024485.1	Scaffold	631.308	LNSY01	396025	0
Asclepias syriaca	Eukaryota; Plants; Land Plants	GCA_002018285.1	Scaffold	236.77	MSXX01	221855	0
Asparagus officinalis	Eukaryota; Plants; Land Plants	GCA_001876935.1	Chromosome	1187.54	MPDI01	11792	36763
Atalantia buxifolia	Eukaryota; Plants; Land Plants	GCA_002013935.1	Scaffold	315.806	MKYR01	25600	0
Aurinia saxatilis	Eukaryota; Plants; Land Plants	GCA_900406295.1	Scaffold	316.42	OVAP01	76972	0
Avena sativa	Eukaryota; Plants; Land Plants	GCA_002943605.1	Contig	67.3266	PKQH01	16667	0
Azadirachta indica	Eukaryota; Plants; Land Plants	GCA_000439995.3	Contig	261.458	AMWY02	126142	0
Barbarea vulgaris	Eukaryota; Plants; Land Plants	GCA_001920985.1	Scaffold	167.352	LXTM01	7810	0
Bassia scoparia	Eukaryota; Plants; Land Plants	GCA_008642245.1	Scaffold	711.357	SNQN01	19671	0
Begonia fuchsioides	Eukaryota; Plants; Land Plants	GCA_003255005.1	Scaffold	373.914	QAOC01	55006	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Berberis thunbergii	Eukaryota; Plants; Land Plants	GCA_003290165.1	Contig	2240.74	QNQO01	11815	0
Beta patula	Eukaryota; Plants; Land Plants	GCA_005862465.1	Scaffold	633.549	VASJ01	78458	0
Beta vulgaris subsp. maritima	Eukaryota; Plants; Land Plants	GCA_005862445.2	Scaffold	608.27	VASK02	97415	0
Beta vulgaris subsp. vulgaris	Eukaryota; Plants; Land Plants	GCA_002917755.1	Chromosome	540.534	PCNB01	40	0
Beta vulgaris subsp. vulgaris	Eukaryota; Plants; Land Plants	GCA_000510975.1	Chromosome	568.609	AYZY01	43635	0
Beta vulgaris subsp. vulgaris	Eukaryota; Plants; Land Plants	GCA_000510365.1	Scaffold	484.231	AYZT01	35771	0
Beta vulgaris subsp. vulgaris	Eukaryota; Plants; Land Plants	GCA_000510875.1	Scaffold	479.876	AYZX01	47405	0
Beta vulgaris subsp. vulgaris	Eukaryota; Plants; Land Plants	GCA_000510485.1	Scaffold	463.706	AYZW01	48733	0
Beta vulgaris subsp. vulgaris	Eukaryota; Plants; Land Plants	GCA_000510465.1	Scaffold	539.552	AYZU01	84234	0
Beta vulgaris subsp. vulgaris	Eukaryota; Plants; Land Plants	GCA_000397105.1	Scaffold	426.675	ARYA01	260142	0
Beta vulgaris subsp. vulgaris	Eukaryota; Plants; Land Plants	GCA_000729925.1	Contig	1.15347	JMBQ01	1287	0
Beta vulgaris subsp. vulgaris	Eukaryota; Plants; Land Plants	GCA_000511025.2	Chromosome	566.55	AYZS02	40406	32874
Betula nana	Eukaryota; Plants; Land Plants	GCA_000327005.1	Scaffold	564.011	CAOK01	551915	0
Betula pendula	Eukaryota; Plants; Land Plants	GCA_900184695.1	Scaffold	435.915	FXXK01	5644	0

Biscutella auriculata	Eukaryota; Plants; Land Plants	GCA_900406285.1	Scaffold	384.77	OVAO01	150640	0
Biscutella laevigata subsp. laevigata	Eukaryota; Plants; Land Plants	GCA_900406315.1	Scaffold	333.609	OVAU01	144912	0
Boechera puberula	Eukaryota; Plants; Land Plants	GCA_900406335.1	Scaffold	182.592	OVAS01	123951	0
Boechera stricta	Eukaryota; Plants; Land Plants	GCA_002079875.1	Scaffold	188.795	MLHT01	1944	0
Boehmeria nivea	Eukaryota; Plants; Land Plants	GCA_002937015.1	Scaffold	344.617	PHNS01	12775	0
Boehmeria nivea	Eukaryota; Plants; Land Plants	GCA_002806895.1	Scaffold	316.026	NHTU01	154955	0
Brachypodium distachyon	Eukaryota; Plants; Land Plants	GCA_002892335.1	Contig	218.676	MXPZ01	60421	0
Brachypodium distachyon	Eukaryota; Plants; Land Plants	GCA_002892295.1	Contig	218.015	MXQA01	65964	0
Brachypodium distachyon	Eukaryota; Plants; Land Plants	GCA_001742125.1	Contig	214.716	LXJM01	68977	0
Brachypodium distachyon	Eukaryota; Plants; Land Plants	GCA_000005505.4	Chromosome	271.299	ADDN03	15	37892
Brassica cretica	Eukaryota; Plants; Land Plants	GCA_003260655.1	Contig	412.521	QGKV01	243461	0
Brassica cretica	Eukaryota; Plants; Land Plants	GCA_003260635.1	Contig	208.354	QGKW01	100644	0
Brassica cretica	Eukaryota; Plants; Land Plants	GCA_003260675.1	Contig	434.935	QGKX01	338759	0
Brassica cretica	Eukaryota; Plants; Land Plants	GCA_003260695.1	Contig	400.212	QGKY01	396633	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Brassica juncea var. tumida	Eukaryota; Plants; Land Plants	GCA_001687265.1	Chromosome	954.861	LFQT01	9746	0
Brassica napus	Eukaryota; Plants; Land Plants	GCA_000751015.1	Scaffold	848.2	CCCW01	20899	61153
Brassica napus	Eukaryota; Plants; Land Plants	GCA_000686985.2	Chromosome	976.191	JMKK02	1471	123467
Brassica nigra	Eukaryota; Plants; Land Plants	GCA_001682895.1	Chromosome	402.145	LFLV01	2545	0
Brassica oleracea	Eukaryota; Plants; Land Plants	GCA_900416815.2	Chromosome	554.977	OWNI02	129	0
Brassica oleracea var. capitata	Eukaryota; Plants; Land Plants	GCA_000604025.1	Scaffold	514.431	AOIX01	1816	0
Brassica oleracea var. oleracea	Eukaryota; Plants; Land Plants	GCA_000695525.1	Chromosome	488.954	JJMF01	32886	56687
Brassica rapa	Eukaryota; Plants; Land Plants	GCA_900412535.2	Chromosome	401.927	OVXL02	304	0
Brassica rapa	Eukaryota; Plants; Land Plants	GCA_000309985.1	Chromosome	284.129	AENI01	40432	52553
Brassica rapa	Eukaryota; Plants; Land Plants	GCA_003434825.1	Chromosome	314.865	QMKI01	7071	43332
Brassica rapa subsp. pekinensis	Eukaryota; Plants; Land Plants	GCA_008629595.1	Chromosome	234.688	VDME01	3421	0
Cajanus cajan	Eukaryota; Plants; Land Plants	GCA_000230855.2	Contig	648.281	AFSP02	360028	0
Cajanus cajan	Eukaryota; Plants; Land Plants	GCA_000340665.1	Chromosome	592.971	AGCT01	36536	41387
Calamus simplicifolius	Eukaryota; Plants; Land Plants	GCA_900491605.1	Scaffold	1960.81	UESW01	5116	0

Calotropis procera	Eukaryota; Plants; Land Plants	GCA_004801955.1	Scaffold	209.215	LVCA01	20519	0
Camelina sativa	Eukaryota; Plants; Land Plants	GCA_000496875.1	Scaffold	547.649	AUUT01	15937	0
Camelina sativa	Eukaryota; Plants; Land Plants	GCA_000633955.1	Chromosome	641.356	JFZQ01	37212	107481
Camellia sinensis var. sinensis	Eukaryota; Plants; Land Plants	GCA_004153795.1	Scaffold	3105.37	SDRB01	14028	76698
Cannabis sativa	Eukaryota; Plants; Land Plants	GCA_001865755.1	Contig	585.824	MNPR01	11110	0
Cannabis sativa	Eukaryota; Plants; Land Plants	GCA_002090435.1	Contig	512.174	MXBD01	18355	0
Cannabis sativa	Eukaryota; Plants; Land Plants	GCA_003660325.2	Contig	1333.38	QVPT02	3372	0
Cannabis sativa	Eukaryota; Plants; Land Plants	GCA_001509995.1	Scaffold	285.933	LKUB01	175088	0
Cannabis sativa	Eukaryota; Plants; Land Plants	GCA_000230575.5	Chromosome	891.965	AGQN03	12836	0
Cannabis sativa	Eukaryota; Plants; Land Plants	GCA_003417725.2	Chromosome	1009.67	QKVJ02	5303	0
Cannabis sativa	Eukaryota; Plants; Land Plants	GCA_900626175.1	Chromosome	876.148	UZAU01	221	33677
Cannabis sativa subsp. indica	Eukaryota; Plants; Land Plants	GCA_001510005.1	Contig	595.358	LKUA01	311039	0
Capsella bursa-pastoris	Eukaryota; Plants; Land Plants	GCA_001974645.1	Scaffold	268.431	MPGU01	8186	0
Capsella rubella	Eukaryota; Plants; Land Plants	GCA_000375325.1	Scaffold	133.064	ANNY01	773	34126

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Capsicum annuum	Eukaryota; Plants; Land Plants	GCA_002878395.2	Chromosome	3212.12	NPHV01	81378	0
Capsicum annuum	Eukaryota; Plants; Land Plants	GCA_000512255.2	Chromosome	3063.86	AYRZ02	35797	35845
Capsicum annuum	Eukaryota; Plants; Land Plants	GCA_000710875.1	Chromosome	2935.88	ASJU01	6478	45410
Capsicum annuum var. glabriusculum	Eukaryota; Plants; Land Plants	GCA_000950795.1	Chromosome	2768.13	ASJV01	16998	0
Capsicum baccatum	Eukaryota; Plants; Land Plants	GCA_002271885.2	Chromosome	3215.61	MLFT02	23260	35853
Capsicum chinense	Eukaryota; Plants; Land Plants	GCA_002271895.2	Chromosome	3070.91	MCIT02	87978	34974
Carica papaya	Eukaryota; Plants; Land Plants	GCA_000150535.1	Scaffold	370.419	ABIM01	17766	26103
Carnegiea gigantea	Eukaryota; Plants; Land Plants	GCA_002740515.1	Scaffold	980.351	NCQR01	57405	0
Carpinus fangiana	Eukaryota; Plants; Land Plants	GCA_006937295.1	Chromosome	381.949	VIBQ01	4602	0
Carthamus tinctorius	Eukaryota; Plants; Land Plants	GCA_001633085.1	Scaffold	661.938	LUCG01	463906	0
Caryocar brasiliense	Eukaryota; Plants; Land Plants	GCA_004918865.1	Scaffold	212.173	STGP01	55248	0
Castanea mollissima	Eukaryota; Plants; Land Plants	GCA_000763605.1	Scaffold	833.241	JRKL01	133589	0
Casuarina equisetifolia subsp. incana	Eukaryota; Plants; Land Plants	GCA_003795335.1	Scaffold	301.458	RDRV01	2936	0
Casuarina glauca	Eukaryota; Plants; Land Plants	GCA_003255045.1	Scaffold	282.811	QAOB01	39787	0

Catharanthus roseus	Eukaryota; Plants; Land Plants	GCA_000949345.1	Scaffold	522.654	JQHZ01	79302	0
Catharanthus roseus	Eukaryota; Plants; Land Plants	GCA_001292525.1	Contig	0.11552	CCXB01	7	0
Catharanthus roseus	Eukaryota; Plants; Land Plants	GCA_001292565.1	Contig	0.114931	CCXA01	32	0
Cenchrus americanus	Eukaryota; Plants; Land Plants	GCA_002174835.2	Chromosome	1816.95	LKME02	52033	0
Cephalotus follicularis	Eukaryota; Plants; Land Plants	GCA_001972305.1	Scaffold	1614.52	BDDD01	16307	36667
Cercis canadensis	Eukaryota; Plants; Land Plants	GCA_003255065.1	Scaffold	329.325	QAOA01	8828	0
Chamaecrista fasciculata	Eukaryota; Plants; Land Plants	GCA_003254925.1	Scaffold	429.103	QANZ01	56674	0
Chenopodium pallidicaule	Eukaryota; Plants; Land Plants	GCA_001687005.1	Scaffold	337.011	MATR01	3013	0
Chenopodium quinoa	Eukaryota; Plants; Land Plants	GCA_001683475.1	Scaffold	1333.55	LPWI01	3487	63173
Chenopodium quinoa	Eukaryota; Plants; Land Plants	GCA_002732095.1	Scaffold	1336.74	NSDK01	3185	0
Chenopodium quinoa	Eukaryota; Plants; Land Plants	GCA_001742885.1	Scaffold	1087.41	BDCQ01	24845	0
Chenopodium suecicum	Eukaryota; Plants; Land Plants	GCA_001687025.1	Scaffold	536.949	MATQ01	11198	0
Chrysanthemum seticuspe	Eukaryota; Plants; Land Plants	GCA_004359105.1	Scaffold	2721.84	BDUE01	354212	0
Cicer arietinum	Eukaryota; Plants; Land Plants	GCA_002896005.2	Scaffold	653.867	PGTT02	13064	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Cicer arietinum	Eukaryota; Plants; Land Plants	GCA_000347275.4	Chromosome	511.684	AHII03	30401	0
Cicer arietinum	Eukaryota; Plants; Land Plants	GCA_000331145.1	Chromosome	530.894	ANPC01	7545	35754
Cicer echinospermum	Eukaryota; Plants; Land Plants	GCA_002896215.2	Scaffold	657.414	PGTU02	17305	0
Cicer reticulatum	Eukaryota; Plants; Land Plants	GCA_002896235.1	Contig	715.407	PGWS01	38802	0
Cicer reticulatum	Eukaryota; Plants; Land Plants	GCA_003689015.2	Chromosome	416.904	QSLP02	3657	0
Cinnamomum micranthum f. kanehirae	Eukaryota; Plants; Land Plants	GCA_003546025.1	Scaffold	730.416	QPKB01	2150	26531
Cissus quadrangularis	Eukaryota; Plants; Land Plants	GCA_002878655.1	Scaffold	281.704	LLYR01	125206	0
Citrullus lanatus	Eukaryota; Plants; Land Plants	GCA_000238415.2	Chromosome	365.45	AGCB02	113	0
Citrus cavaleriei	Eukaryota; Plants; Land Plants	GCA_002013975.2	Scaffold	357.621	MKYP02	14916	0
Citrus clementina	Eukaryota; Plants; Land Plants	GCA_000493195.1	Scaffold	301.365	AMZM01	1398	32586
Citrus hindsii	Eukaryota; Plants; Land Plants	GCA_004802465.1	Contig	373.17	QWBT01	1331	0
Citrus maxima	Eukaryota; Plants; Land Plants	GCA_002006925.1	Chromosome	345.757	MKYQ01	1612	0
Citrus medica	Eukaryota; Plants; Land Plants	GCA_002013955.2	Scaffold	406.058	MKYO02	32732	0
Citrus reticulata	Eukaryota; Plants; Land Plants	GCA_003258625.1	Scaffold	344.273	NIHA01	67725	0

Citrus sinensis	Eukaryota; Plants; Land Plants	GCA_000695605.1	Scaffold	319.225	JJOQ01	12573	51718
Citrus sinensis	Eukaryota; Plants; Land Plants	GCA_000317415.1	Chromosome	327.83	AJPS01	4995	39056
Citrus unshiu	Eukaryota; Plants; Land Plants	GCA_002897195.1	Scaffold	359.652	BDQV01	20876	37970
Citrus unshiu	Eukaryota; Plants; Land Plants	GCA_001753815.1	Contig	1.1542	BDGO01	507	0
Citrus x paradisi x Citrus trifoliata	Eukaryota; Plants; Land Plants	GCA_001929425.1	Contig	265.534	AZHM01	238488	0
Cochlearia officinalis	Eukaryota; Plants; Land Plants	GCA_900406305.1	Scaffold	164.469	OVAT01	128459	0
Cocos nucifera	Eukaryota; Plants; Land Plants	GCA_006176705.1	Scaffold	2102.42	QRFJ01	7998	0
Cocos nucifera	Eukaryota; Plants; Land Plants	GCA_003604295.1	Scaffold	1839.17	PDMH01	59328	0
Cocos nucifera	Eukaryota; Plants; Land Plants	GCA_008124465.1	Chromosome	2202.46	VOII01	113653	0
Codonopsis pilosula	Eukaryota; Plants; Land Plants	GCA_004523855.1	Scaffold	937.71	SOPR01	1154	0
Coffea arabica	Eukaryota; Plants; Land Plants	GCA_003713225.1	Chromosome	1094.45	RHJU01	2833	67222
Coffea canephora	Eukaryota; Plants; Land Plants	GCA_900059795.1	Chromosome	568.612	CBUE02	13345	25574
Coffea eugenioides	Eukaryota; Plants; Land Plants	GCA_003713205.1	Chromosome	699.904	RHJT01	3530	38150
Conringia planisiliqua	Eukaryota; Plants; Land Plants	GCA_900108845.1	Scaffold	184.156	FNXX01	705	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Corchorus capsularis	Eukaryota; Plants; Land Plants	GCA_001974805.1	Scaffold	317.178	AWWV01	16522	29356
Corchorus olitorius	Eukaryota; Plants; Land Plants	GCA_001974825.1	Contig	334.912	AWUE01	24918	35704
Corchorus olitorius	Eukaryota; Plants; Land Plants	GCA_002141455.1	Contig	377.377	LLWS01	52373	0
Crucihimalaya himalaica	Eukaryota; Plants; Land Plants	GCA_004349715.1	Scaffold	234.721	SMJT01	582	0
Cucumis melo	Eukaryota; Plants; Land Plants	GCA_000313045.1	Scaffold	374.928	CAJI01	31464	29798
Cucumis melo	Eukaryota; Plants; Land Plants	GCA_902497455.1	Scaffold	357.857	CABVGI01	13	0
Cucumis melo var. makuwa	Eukaryota; Plants; Land Plants	GCA_005549215.1	Scaffold	358.48	SSTE01	23444	38173
Cucumis melo var. makuwa	Eukaryota; Plants; Land Plants	GCA_005549225.1	Scaffold	347.184	SSTD01	20255	36235
Cucumis sativus	Eukaryota; Plants; Land Plants	GCA_001483825.2	Contig	342.654	LKUO02	8035	0
Cucumis sativus	Eukaryota; Plants; Land Plants	GCA_000224045.1	Scaffold	323.986	ACYN01	13113	0
Cucumis sativus	Eukaryota; Plants; Land Plants	GCA_000004075.2	Chromosome	195.669	ACHR02	190	25668
Cucurbita argyrosperma subsp. argyrosperma	Eukaryota; Plants; Land Plants	GCA_004115005.1	Scaffold	230.034	SDJN01	938	0
Cucurbita maxima	Eukaryota; Plants; Land Plants	GCA_002738345.1	Scaffold	271.413	NEWN01	8299	42777
Cucurbita moschata	Eukaryota; Plants; Land Plants	GCA_002738365.1	Scaffold	269.943	NEWM01	3500	43715

Cucurbita pepo subsp. pepo	Eukaryota; Plants; Land Plants	GCA_002806865.2	Chromosome	261.355	NHTM01	25263	43466
Cuscuta australis	Eukaryota; Plants; Land Plants	GCA_003260385.1	Contig	262.63	NQVE01	218	18157
Cuscuta campestris	Eukaryota; Plants; Land Plants	GCA_900332095.1	Scaffold	476.792	OOIL01	6907	0
Cynara cardunculus var. scolymus	Eukaryota; Plants; Land Plants	GCA_001531365.1	Chromosome	725.198	LEKV01	13588	38406
Dactylis glomerata	Eukaryota; Plants; Land Plants	GCA_007115705.1	Scaffold	1781.32	QXEO01	2117	0
Dactylis glomerata	Eukaryota; Plants; Land Plants	GCA_002892645.1	Scaffold	839.915	MVYT01	1072009	0
Datisca glomerata	Eukaryota; Plants; Land Plants	GCA_003255025.1	Scaffold	688.404	QANY01	13864	0
Daucus carota subsp. sativus	Eukaryota; Plants; Land Plants	GCA_001625215.1	Chromosome	421.539	LNRQ01	4826	44655
Dendrobium catenatum	Eukaryota; Plants; Land Plants	GCA_001605985.2	Scaffold	1104.26	JSDN02	286090	34389
Dianthus caryophyllus	Eukaryota; Plants; Land Plants	GCA_000512335.1	Scaffold	567.662	BAUD01	45088	0
Dichanthelium oligosanthes	Eukaryota; Plants; Land Plants	GCA_001633215.2	Scaffold	589.166	LWDX02	17436	26468
Dioscorea alata	Eukaryota; Plants; Land Plants	GCA_002904275.2	Scaffold	620.909	CZHE02	57706	0
Dioscorea rotundata	Eukaryota; Plants; Land Plants	GCA_002260605.1	Scaffold	594.227	BBQW01	4723	0
Dioscorea rotundata	Eukaryota; Plants; Land Plants	GCA_002260665.1	Scaffold	730.21	BDML01	615107	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Dioscorea rotundata	Eukaryota; Plants; Land Plants	GCA_002260645.1	Scaffold	683.283	BDMK01	641416	0
Dioscorea rotundata	Eukaryota; Plants; Land Plants	GCA_002240015.2	Chromosome	456.675	BDMI01	21	0
Dioscorea sansibarensis	Eukaryota; Plants; Land Plants	GCA_900631875.1	Contig	0.128321	CABFPD01	6	0
Diospyros lotus	Eukaryota; Plants; Land Plants	GCA_000774125.1	Scaffold	1.10419	JRBH01	796	0
Dorcoceras hygrometricum	Eukaryota; Plants; Land Plants	GCA_001598015.1	Scaffold	1521.36	LVEL01	401752	47778
Drosera capensis	Eukaryota; Plants; Land Plants	GCA_001925005.1	Scaffold	263.788	LIEC01	12713	0
Dryas drummondii	Eukaryota; Plants; Land Plants	GCA_003254865.1	Scaffold	225.547	QANW01	13357	0
Durio zibethinus	Eukaryota; Plants; Land Plants	GCA_002303985.1	Scaffold	715.23	NSDW01	677	63007
Echinochloa crus-galli	Eukaryota; Plants; Land Plants	GCA_900205405.1	Scaffold	1486.61	OAMR01	4534	0
Echium plantagineum	Eukaryota; Plants; Land Plants	GCA_003412495.2	Chromosome	349.028	QFAX02	809	0
Eichhornia paniculata	Eukaryota; Plants; Land Plants	GCA_001647135.1	Scaffold	571.388	LTAE01	40286	0
Elaeis guineensis	Eukaryota; Plants; Land Plants	GCA_001672495.1	Scaffold	499.029	JRVM01	218141	0
Elaeis guineensis	Eukaryota; Plants; Land Plants	GCA_002146295.1	Contig	134.97	AXCU01	186862	0
Elaeis guineensis	Eukaryota; Plants; Land Plants	GCA_000442705.1	Chromosome	1535.18	ASJS01	40349	43551

Elaeis oleifera	Eukaryota; Plants; Land Plants	GCA_000441515.1	Scaffold	1402.73	ASIR01	26756	0
Elaeis oleifera	Eukaryota; Plants; Land Plants	GCA_002146275.1	Contig	60.0019	AXCH01	93897	0
Eleusine coracana subsp. coracana	Eukaryota; Plants; Land Plants	GCA_002180455.1	Scaffold	1195.99	LXGH01	525627	0
Eleusine indica	Eukaryota; Plants; Land Plants	GCA_003369855.1	Scaffold	492.27	QEPD01	24072	0
Embelia ribes	Eukaryota; Plants; Land Plants	GCA_001753735.1	Scaffold	660.51	MKEJ01	107000	0
Ensete ventricosum	Eukaryota; Plants; Land Plants	GCA_000818735.3	Scaffold	451.279	JTFG03	45742	58438
Ensete ventricosum	Eukaryota; Plants; Land Plants	GCA_001884845.1	Scaffold	444.842	MKKT01	51525	58998
Ensete ventricosum	Eukaryota; Plants; Land Plants	GCA_001884805.1	Contig	429.48	MKKS01	60129	56086
Ensete ventricosum	Eukaryota; Plants; Land Plants	GCA_000331365.3	Scaffold	437.269	AMZH03	52691	55115
Eragrostis curvula	Eukaryota; Plants; Land Plants	GCA_007726485.1	Chromosome	603.072	RWGY01	1143	55182
Eragrostis tef	Eukaryota; Plants; Land Plants	GCA_000970635.1	Scaffold	607.318	LAPY01	13883	0
Erigeron canadensis	Eukaryota; Plants; Land Plants	GCA_000775935.1	Contig	326.165	JSWR01	20075	0
Erucastrum elatum	Eukaryota; Plants; Land Plants	GCA_900406325.1	Scaffold	362.114	OVBX01	80289	0
Erysimum cheiri	Eukaryota; Plants; Land Plants	GCA_900406345.1	Scaffold	147.321	OVAQ01	37531	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Erysimum pusillum	Eukaryota; Plants; Land Plants	GCA_900406355.1	Scaffold	184.064	OVBW01	47441	0
Erythranthe guttata	Eukaryota; Plants; Land Plants	GCA_000504015.1	Scaffold	322.167	APLE01	2212	31861
Eschscholzia californica subsp. californica	Eukaryota; Plants; Land Plants	GCA_002897215.1	Scaffold	489.065	BEHA01	53253	0
Eucalyptus camaldulensis	Eukaryota; Plants; Land Plants	GCA_000260855.1	Contig	654.922	BADO01	274001	0
Eucalyptus grandis	Eukaryota; Plants; Land Plants	GCA_000612305.1	Scaffold	691.43	AUSX01	4951	52554
Eucalyptus melliodora	Eukaryota; Plants; Land Plants	GCA_004368105.1	Scaffold	643.228	SISH01	423	0
Eucalyptus pauciflora	Eukaryota; Plants; Land Plants	GCA_007663325.1	Scaffold	594.528	VMYD01	415	0
Euclidium syriacum	Eukaryota; Plants; Land Plants	GCA_900116095.1	Scaffold	229.211	FPAK01	160	0
Eugenia uniflora	Eukaryota; Plants; Land Plants	GCA_004012085.1	Contig	3.15213	RQIG01	2601	0
Euphorbia esula	Eukaryota; Plants; Land Plants	GCA_002919075.1	Scaffold	1124.89	PJAD01	1633094	0
Euphorbia esula	Eukaryota; Plants; Land Plants	GCA_002918425.1	Scaffold	639.02	PJAE01	912031	0
Eutrema heterophyllum	Eukaryota; Plants; Land Plants	GCA_002933915.1	Scaffold	348.971	PKMM01	57686	0
Eutrema salsugineum	Eukaryota; Plants; Land Plants	GCA_000478725.1	Scaffold	243.11	ANOA01	638	33637
Eutrema salsugineum	Eukaryota; Plants; Land Plants	GCA_000325905.2	Chromosome	231.893	AHIU01	2663	0

Eutrema yunnanense	Eukaryota; Plants; Land Plants	GCA_002933935.1	Scaffold	415.364	PKML01	78020	0
Fagopyrum esculentum	Eukaryota; Plants; Land Plants	GCA_004303065.1	Scaffold	1087.7	QWEW01	14903	0
Fagopyrum esculentum	Eukaryota; Plants; Land Plants	GCA_001661195.1	Scaffold	1177.69	BCYN01	387594	0
Fagopyrum tataricum	Eukaryota; Plants; Land Plants	GCA_002928575.1	Scaffold	526.768	PKMW01	2564	0
Fagopyrum tataricum	Eukaryota; Plants; Land Plants	GCA_002319775.1	Chromosome	505.883	NCTC01	7020	0
Fagus sylvatica	Eukaryota; Plants; Land Plants	GCA_003347535.1	Scaffold	428.2	QCXR01	8673	0
Ficus carica	Eukaryota; Plants; Land Plants	GCA_002002945.1	Scaffold	247.091	BDEM01	27995	0
Ficus erecta	Eukaryota; Plants; Land Plants	GCA_008635985.1	Contig	595.835	BKCH01	2455	0
Foeniculum vulgare	Eukaryota; Plants; Land Plants	GCA_003724115.1	Scaffold	1010.97	PHNY01	300377	0
Fragaria iinumae	Eukaryota; Plants; Land Plants	GCA_000511975.1	Scaffold	199.628	BATU01	117822	0
Fragaria nipponica	Eukaryota; Plants; Land Plants	GCA_000512025.1	Scaffold	206.415	BATV01	215024	0
Fragaria nubicola	Eukaryota; Plants; Land Plants	GCA_000511995.1	Scaffold	203.686	BATW01	210780	0
Fragaria orientalis	Eukaryota; Plants; Land Plants	GCA_000517285.1	Scaffold	214.184	BATX01	323163	0
Fragaria vesca subsp. vesca	Eukaryota; Plants; Land Plants	GCA_000184155.1	Chromosome	214.373	AEMH01	3048	31387

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Fragaria x ananassa	Eukaryota; Plants; Land Plants	GCA_000511835.1	Scaffold	697.762	BATT01	625966	0
Fragaria x ananassa	Eukaryota; Plants; Land Plants	GCA_000511695.1	Scaffold	173.23	BATS01	211588	0
Fraxinus excelsior	Eukaryota; Plants; Land Plants	GCA_900149125.1	Scaffold	867.455	FTPI01	89515	0
Gastrodiaelata f. glauca	Eukaryota; Plants; Land Plants	GCA_002966915.1	Scaffold	1060.98	PVEL01	3768	0
Genlisea aurea	Eukaryota; Plants; Land Plants	GCA_000441915.1	Scaffold	43.3578	AUSU01	10684	17685
Geum urbanum	Eukaryota; Plants; Land Plants	GCA_900236755.1	Scaffold	1217.04	OEJZ01	170029	0
Glycine max	Eukaryota; Plants; Land Plants	GCA_001269945.2	Contig	927.706	BBNX02	108601	0
Glycine max	Eukaryota; Plants; Land Plants	GCA_003349995.1	Chromosome	1017.57	QKRT01	495	0
Glycine max	Eukaryota; Plants; Land Plants	GCA_002905335.2	Chromosome	1016.28	PELE01	475	0
Glycine max	Eukaryota; Plants; Land Plants	GCA_000004515.4	Chromosome	979.046	ACUP03	1579	71219
Glycine soja	Eukaryota; Plants; Land Plants	GCA_000722935.2	Scaffold	863.568	AZNC01	33170	50399
Glycine soja	Eukaryota; Plants; Land Plants	GCA_004193775.2	Chromosome	1013.77	QZWG01	1120	69277
Glycine soja	Eukaryota; Plants; Land Plants	GCA_002907465.1	Chromosome	985.26	PGFP01	805	0
Glycine tomentella	Eukaryota; Plants; Land Plants	GCA_007407185.1	Scaffold	1694.09	PYAF01	6353	0

Gossypioides kirkii	Eukaryota; Plants; Land Plants	GCA_002818315.1	Chromosome	528.715	PEQG01	745	0
Gossypium arboreum	Eukaryota; Plants; Land Plants	GCA_000787975.1	Scaffold	1862.24	JRRC01	392831	33609
Gossypium arboreum	Eukaryota; Plants; Land Plants	GCA_000612285.2	Chromosome	1694.6	AYOE01	75419	47568
Gossypium australe	Eukaryota; Plants; Land Plants	GCA_005393395.2	Chromosome	1743.39	SMMG02	564	38281
Gossypium barbadense	Eukaryota; Plants; Land Plants	GCA_002928715.1	Scaffold	1394.24	LAGA01	9269	40359
Gossypium barbadense	Eukaryota; Plants; Land Plants	GCA_001856525.1	Scaffold	2566.74	AXCG01	29751	0
Gossypium barbadense	Eukaryota; Plants; Land Plants	GCA_002926015.1	Scaffold	775.252	LAGB01	4265	36871
Gossypium barbadense	Eukaryota; Plants; Land Plants	GCA_008761655.1	Chromosome	2195.8	VKDL01	4748	108363
Gossypium darwinii	Eukaryota; Plants; Land Plants	GCA_007990325.1	Chromosome	2182.96	VKGI01	821	97407
Gossypium hirsutum	Eukaryota; Plants; Land Plants	GCA_006980745.1	Chromosome	2287.87	VCQY01	599	0
Gossypium hirsutum	Eukaryota; Plants; Land Plants	GCA_006980775.1	Chromosome	2308.22	VCQX01	2238	0
Gossypium hirsutum	Eukaryota; Plants; Land Plants	GCA_000987745.1	Chromosome	2189.14	LBLM01	9148	90927
Gossypium mustelinum	Eukaryota; Plants; Land Plants	GCA_007990455.1	Chromosome	2315.09	VKGF01	2146	106487
Gossypium raimondii	Eukaryota; Plants; Land Plants	GCA_000331045.1	Scaffold	773.768	AMOP01	4699	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Gossypium raimondii	Eukaryota; Plants; Land Plants	GCA_000327365.1	Chromosome	761.565	ALYE01	1034	59057
Gossypium thurberi	Eukaryota; Plants; Land Plants	GCA_004027125.1	Chromosome	582.007	RCOT01	15297	0
Gossypium tomentosum	Eukaryota; Plants; Land Plants	GCA_007990485.1	Chromosome	2193.56	VKGE01	749	112713
Handroanthus impetiginosus	Eukaryota; Plants; Land Plants	GCA_002762385.1	Scaffold	503.289	NKXS01	13204	30271
Helianthus annuus	Eukaryota; Plants; Land Plants	GCA_002127325.1	Chromosome	3027.84	MNCJ01	1528	73839
Heliophila coronopifolia	Eukaryota; Plants; Land Plants	GCA_900406365.1	Scaffold	293.602	OVAW01	212649	0
Herrania umbratica	Eukaryota; Plants; Land Plants	GCA_002168275.2	Scaffold	234.039	NHTG01	6074	27748
Hevea brasiliensis	Eukaryota; Plants; Land Plants	GCA_001654055.1	Scaffold	1373.53	LVXX01	7453	58062
Hevea brasiliensis	Eukaryota; Plants; Land Plants	GCA_002003025.1	Scaffold	1256.27	BDHL01	592579	0
Hevea brasiliensis	Eukaryota; Plants; Land Plants	GCA_001907995.1	Scaffold	1550.51	MKXE01	189320	0
Hevea brasiliensis	Eukaryota; Plants; Land Plants	GCA_000340545.1	Scaffold	1301.4	AJJZ01	1150326	0
Hibiscus syriacus	Eukaryota; Plants; Land Plants	GCA_006381635.1	Scaffold	2573.67	VEPZ01	9646	0
Hibiscus syriacus	Eukaryota; Plants; Land Plants	GCA_001696755.1	Scaffold	1748.25	MBGJ01	77488	0
Hordeum bulbosum	Eukaryota; Plants; Land Plants	GCA_900070015.1	Scaffold	1294.87	CBQS01	2883554	0

Hordeum pubiflorum	Eukaryota; Plants; Land Plants	GCA_000582825.1	Scaffold	1425.27	CBMN01	1818420	0
Hordeum vulgare	Eukaryota; Plants; Land Plants	GCA_004114815.1	Scaffold	4006.12	SDOW01	1856	0
Hordeum vulgare	Eukaryota; Plants; Land Plants	GCA_900075435.2	Scaffold	9788.86	FJWB02	72295	0
Hordeum vulgare	Eukaryota; Plants; Land Plants	GCA_900067795.1	Contig	161.166	CEGI01	589134	0
Hordeum vulgare	Eukaryota; Plants; Land Plants	GCA_900067805.1	Contig	160.635	CEGJ01	598302	0
Hordeum vulgare	Eukaryota; Plants; Land Plants	GCA_900067785.1	Contig	316.11	CEGM01	1201146	0
Hordeum vulgare	Eukaryota; Plants; Land Plants	GCA_900067825.1	Contig	207.795	CEGK01	800872	0
Hordeum vulgare	Eukaryota; Plants; Land Plants	GCA_900067815.1	Contig	233.129	CEGL01	908607	0
Hordeum vulgare	Eukaryota; Plants; Land Plants	GCA_000947855.1	Contig	98.056	CEGH01	391258	0
Hordeum vulgare subsp. vulgare	Eukaryota; Plants; Land Plants	GCA_000326125.1	Scaffold	1779.49	CAJX01	2077901	0
Hordeum vulgare subsp. vulgare	Eukaryota; Plants; Land Plants	GCA_001077415.1	Scaffold	1645.58	CBLZ01	2280908	0
Hordeum vulgare subsp. vulgare	Eukaryota; Plants; Land Plants	GCA_000326085.1	Scaffold	1868.64	CAJW01	2670738	0
Hordeum vulgare subsp. vulgare	Eukaryota; Plants; Land Plants	GCA_900002345.1	Scaffold	1825.17	CCJR01	2546226	0
Hordeum vulgare subsp. vulgare	Eukaryota; Plants; Land Plants	GCA_002900805.1	Scaffold	2019.37	CAJV01	2742077	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Hordeum vulgare subsp. vulgare	Eukaryota; Plants; Land Plants	GCA_902500625.1	Scaffold	4129.36	CABVVH01	8	0
Hordeum vulgare subsp. vulgare	Eukaryota; Plants; Land Plants	GCA_000227425.1	Contig	28.016	BACC01	8583	0
Hordeum vulgare subsp. vulgare	Eukaryota; Plants; Land Plants	GCA_002943585.1	Contig	57.9644	PKQG01	14158	0
Humulus lupulus var. cordifolius	Eukaryota; Plants; Land Plants	GCA_000830395.1	Scaffold	2049.21	BBPB01	132476	0
Humulus lupulus var. lupulus	Eukaryota; Plants; Land Plants	GCA_000831365.1	Scaffold	2049.21	BBPC01	132476	0
Iberis amara	Eukaryota; Plants; Land Plants	GCA_900406375.1	Scaffold	361.245	OVAV01	191171	0
Iberis pinnata	Eukaryota; Plants; Land Plants	GCA_900406425.1	Scaffold	686.283	OVBE01	141542	0
Ipomoea batatas	Eukaryota; Plants; Land Plants	GCA_900092185.1	Contig	13.8613	FLTB01	41487	0
Ipomoea batatas	Eukaryota; Plants; Land Plants	GCA_002525835.2	Chromosome	837.013	NXFB01	28461	0
Ipomoea nil	Eukaryota; Plants; Land Plants	GCA_001879475.1	Scaffold	735.231	BDFN01	3418	51054
Ipomoea trifida	Eukaryota; Plants; Land Plants	GCA_000978395.1	Scaffold	512.991	BBOG01	77400	0
Ipomoea trifida	Eukaryota; Plants; Land Plants	GCA_000981105.1	Scaffold	712.155	BBOH01	181194	0
Ipomoea trifida	Eukaryota; Plants; Land Plants	GCA_004706985.1	Chromosome	460.934	SMMV01	4236	0
Isatis lusitanica	Eukaryota; Plants; Land Plants	GCA_900406415.1	Scaffold	203.919	OVBB01	61090	0

Isatis tinctoria	Eukaryota; Plants; Land Plants	GCA_900406385.1	Scaffold	244.177	OVBD01	106900	0
Jaltoma tasinuosa	Eukaryota; Plants; Land Plants	GCA_003996215.1	Scaffold	1443.2	QJPP01	7667	0
Jatropha curcas	Eukaryota; Plants; Land Plants	GCA_000696525.1	Scaffold	318.527	AFEW01	6024	32547
Jatropha curcas	Eukaryota; Plants; Land Plants	GCA_004143595.1	Scaffold	265.767	QFZG01	2959	0
Jatropha curcas	Eukaryota; Plants; Land Plants	GCA_000208675.2	Scaffold	297.661	BABX02	39277	0
Juglans cathayensis	Eukaryota; Plants; Land Plants	GCA_003122765.1	Scaffold	600.151	QEOU01	19972	0
Juglans hindsii	Eukaryota; Plants; Land Plants	GCA_003123825.1	Scaffold	611.109	QEOW01	73433	0
Juglans mandshurica	Eukaryota; Plants; Land Plants	GCA_002916435.1	Scaffold	558.071	PKSJ01	13809	0
Juglans microcarpa	Eukaryota; Plants; Land Plants	GCA_003123845.1	Scaffold	913.972	QEOX01	112570	0
Juglans microcarpa x Juglans regia	Eukaryota; Plants; Land Plants	GCA_004785585.1	Chromosome	534.672	QKZY01	73	0
Juglans microcarpa x Juglans regia	Eukaryota; Plants; Land Plants	GCA_004785595.1	Chromosome	527.896	QKZX01	154	0
Juglans nigra	Eukaryota; Plants; Land Plants	GCA_003123865.1	Scaffold	620.767	QEOV01	90472	0
Juglans nigra	Eukaryota; Plants; Land Plants	GCA_002916485.1	Scaffold	682.557	PKSI01	18575	0
Juglans regia	Eukaryota; Plants; Land Plants	GCA_001411555.1	Scaffold	699.673	LIHL01	105803	55627

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Juglans regia	Eukaryota; Plants; Land Plants	GCA_003122785.1	Scaffold	650.478	QEOZ01	4401	0
Juglans regia	Eukaryota; Plants; Land Plants	GCA_002916465.1	Scaffold	634.748	PKSH01	25789	0
Juglans sigillata	Eukaryota; Plants; Land Plants	GCA_003123805.1	Scaffold	648.117	QEOY01	134300	0
Kalanchoe fedtschenkoi	Eukaryota; Plants; Land Plants	GCA_002312845.1	Scaffold	256.351	NQLW01	1324	0
Kernera saxatilis	Eukaryota; Plants; Land Plants	GCA_900406395.1	Scaffold	143.969	OVAY01	18372	0
Kokia drynarioides	Eukaryota; Plants; Land Plants	GCA_002814295.1	Scaffold	517.43	NTFQ01	15383	0
Lactuca sativa	Eukaryota; Plants; Land Plants	GCA_002870075.1	Scaffold	2384.19	NBSK01	11453	45242
Lactuca sativa	Eukaryota; Plants; Land Plants	GCA_900243165.1	Scaffold	2224.43	OFAD01	161898	0
Lactuca sativa	Eukaryota; Plants; Land Plants	GCA_900198505.1	Scaffold	1975.25	FZNH01	138326	0
Lactuca sativa	Eukaryota; Plants; Land Plants	GCA_000227445.1	Contig	1133.66	AFSA01	876110	0
Lagenaria siceraria	Eukaryota; Plants; Land Plants	GCA_003268545.1	Scaffold	313.387	NHZF01	438	0
Lagenaria siceraria	Eukaryota; Plants; Land Plants	GCA_000466325.1	Scaffold	176.727	ATBX01	305112	0
Lagenaria siceraria	Eukaryota; Plants; Land Plants	GCA_002890555.2	Chromosome	297.879	MIMD02	27	0
Larix sibirica	Eukaryota; Plants; Land Plants	GCA_004151065.1	Scaffold	12342.1	NWUY01	11325800	0

Leavenworthiaalabamica	Eukaryota; Plants; Land Plants	GCA_000411055.1	Scaffold	173.432	ASXC01	11715	0
Leersia perrieri	Eukaryota; Plants; Land Plants	GCA_000325765.3	Chromosome	266.688	ALNV02	12	0
Lepidium africanum	Eukaryota; Plants; Land Plants	GCA_900406405.1	Scaffold	225.262	OVAX01	17445	0
Lepidium aucheri	Eukaryota; Plants; Land Plants	GCA_900406435.1	Scaffold	332.063	OVBA01	31191	0
Lindernia brevidens	Eukaryota; Plants; Land Plants	GCA_004919715.1	Scaffold	266.105	SWDC01	158	0
Liriodendron chinense	Eukaryota; Plants; Land Plants	GCA_003013855.2	Scaffold	1742.42	PVNU02	3710	0
Lolium perenne	Eukaryota; Plants; Land Plants	GCA_001735685.1	Scaffold	481.479	MEHO01	666180	0
Lophocereus schottii	Eukaryota; Plants; Land Plants	GCA_002740545.1	Scaffold	797.926	NCQV01	158704	0
Lotus japonicus	Eukaryota; Plants; Land Plants	GCA_000181115.2	Contig	394.455	BABK02	44464	0
Lupinus angustifolius	Eukaryota; Plants; Land Plants	GCA_000338175.1	Scaffold	523.298	AOCW01	71995	0
Lupinus angustifolius	Eukaryota; Plants; Land Plants	GCA_001865875.1	Chromosome	609.203	MLAU01	14378	52821
Macadamia integrifolia	Eukaryota; Plants; Land Plants	GCA_900631585.1	Scaffold	744.636	UZVR01	4098	0
Macadamia integrifolia	Eukaryota; Plants; Land Plants	GCA_900087525.1	Scaffold	518.49	FLKO01	193493	0
Macleaya cordata	Eukaryota; Plants; Land Plants	GCA_002174775.1	Scaffold	377.834	MVGT01	4547	21911

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Macropodium nivale	Eukaryota; Plants; Land Plants	GCA_900406455.1	Scaffold	300.735	OVBF01	122604	0
Magnolia ashei	Eukaryota; Plants; Land Plants	GCA_003571905.1	Scaffold	284.512	PCNC01	265493	0
Malus baccata	Eukaryota; Plants; Land Plants	GCA_006547085.1	Scaffold	674.412	VIEB01	47473	45900
Malus domestica	Eukaryota; Plants; Land Plants	GCA_004115385.1	Chromosome	660.463	RDQH01	343	42841
Malus domestica	Eukaryota; Plants; Land Plants	GCA_000148765.2	Chromosome	1874.77	ACYM01	1667	0
Malus domestica	Eukaryota; Plants; Land Plants	GCA_002114115.1	Chromosome	703.358	MJAX01	807	52039
Manihot esculenta	Eukaryota; Plants; Land Plants	GCA_000737115.1	Scaffold	292.098	JPQF01	65771	0
Manihot esculenta	Eukaryota; Plants; Land Plants	GCA_003957885.1	Scaffold	1276.89	RSFS01	4440	0
Manihot esculenta	Eukaryota; Plants; Land Plants	GCA_003957995.1	Scaffold	1224.64	RSFT01	5398	0
Manihot esculenta	Eukaryota; Plants; Land Plants	GCA_001659605.1	Chromosome	582.279	LTYI01	2020	43286
Manihot esculenta subsp. flabellifolia	Eukaryota; Plants; Land Plants	GCA_000737105.1	Scaffold	390.836	JPQE01	54016	0
Marchantia inflexa	Eukaryota; Plants; Land Plants	GCA_006177815.1	Scaffold	208.753	QLSQ01	41556	0
Marchantia polymorpha	Eukaryota; Plants; Land Plants	GCA_003032435.1	Scaffold	225.761	PNPG01	2957	24674
Marchantia polymorpha subsp. ruderalis	Eukaryota; Plants; Land Plants	GCA_001641455.1	Scaffold	205.718	LVLJ01	4137	17956

Medicago truncatula	Eukaryota; Plants; Land Plants	GCA_003473485.2	Chromosome	429.612	PSQE01	40	44450
Medicago truncatula	Eukaryota; Plants; Land Plants	GCA_002024945.1	Scaffold	402.065	MWMB01	909	0
Medicago truncatula	Eukaryota; Plants; Land Plants	GCA_002251925.1	Scaffold	427.447	MLKM01	2368	0
Medicago truncatula	Eukaryota; Plants; Land Plants	GCA_002251935.1	Scaffold	394.059	MKZU01	1490	0
Medicago truncatula	Eukaryota; Plants; Land Plants	GCA_002251955.1	Scaffold	426.024	MNAG01	2700	0
Medicago truncatula	Eukaryota; Plants; Land Plants	GCA_000219495.2	Chromosome	412.924	APNO01	2187	41939
Mentha longifolia	Eukaryota; Plants; Land Plants	GCA_001642375.1	Scaffold	353.287	LSBG01	190876	0
Metrosideros polymorpha var. glaberrima	Eukaryota; Plants; Land Plants	GCA_001662345.1	Scaffold	304.366	BCNH01	36376	0
Mimosa pudica	Eukaryota; Plants; Land Plants	GCA_003254945.1	Scaffold	557.202	QANV01	97892	0
Miscanthus sacchariflorus	Eukaryota; Plants; Land Plants	GCA_002993905.1	Chromosome	2074.92	PUID01	137916	0
Momordica charantia	Eukaryota; Plants; Land Plants	GCA_001995035.1	Scaffold	285.614	BDCS01	1052	28666
Momordica charantia	Eukaryota; Plants; Land Plants	GCA_900491585.1	Scaffold	296.263	UESV01	3101	0
Monotropa hypopitys	Eukaryota; Plants; Land Plants	GCA_002855965.1	Contig	2197.49	NMUG01	1259264	0
Morella rubra	Eukaryota; Plants; Land Plants	GCA_003952965.1	Chromosome	313.02	RXIC01	500	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Morusn otabilis	Eukaryota; Plants; Land Plants	GCA_000414095.2	Scaffold	320.379	ATGF01	31301	27648
Mucuna pruriens	Eukaryota; Plants; Land Plants	GCA_003370565.1	Scaffold	397.042	QJKJ01	18487	56019
Musa acuminata subsp. malaccensis	Eukaryota; Plants; Land Plants	GCA_000313855.2	Chromosome	472.231	CAIC01	7512	47707
Musa balbisiana	Eukaryota; Plants; Land Plants	GCA_004837865.1	Chromosome	492.775	PYDT01	2590	33021
Musa itinerans	Eukaryota; Plants; Land Plants	GCA_001649415.1	Scaffold	455.349	LVTN01	28415	0
Musa schizocarpa	Eukaryota; Plants; Land Plants	GCA_900464855.1	Scaffold	525.283	UBIG01	194	0
Nasturtium officinale	Eukaryota; Plants; Land Plants	GCA_900406445.1	Scaffold	216.122	OVAZ01	10793	0
Nelumbo nucifera	Eukaryota; Plants; Land Plants	GCA_000365185.2	Scaffold	804.648	AQOG01	3603	38191
Nelumbo nucifera	Eukaryota; Plants; Land Plants	GCA_000805495.1	Scaffold	790.339	APLB01	14895	0
Nelumbo nucifera	Eukaryota; Plants; Land Plants	GCA_003033685.1	Chromosome	817.268	DLUB01	2341	0
Nelumbo nucifera	Eukaryota; Plants; Land Plants	GCA_003033695.1	Chromosome	799.479	DLUA01	12643	0
Nicotiana attenuata	Eukaryota; Plants; Land Plants	GCA_002018495.1	Scaffold	1827.78	MCOF01	951503	0
Nicotiana attenuata	Eukaryota; Plants; Land Plants	GCA_001879085.1	Chromosome	2365.68	MJEQ01	37194	44491
Nicotiana benthamiana	Eukaryota; Plants; Land Plants	GCA_000723945.1	Contig	61.9511	CBMM01	100480	0

Nicotiana glauca	Eukaryota; Plants; Land Plants	GCA_002930595.1	Scaffold	3222.83	PGPE01	514289	0
Nicotiana knightiana	Eukaryota; Plants; Land Plants	GCA_005239525.1	Scaffold	2298.94	MDKJ01	160415	0
Nicotiana obtusifolia	Eukaryota; Plants; Land Plants	GCA_002018475.1	Scaffold	1222.77	MCJB01	53128	0
Nicotiana otophora	Eukaryota; Plants; Land Plants	GCA_000715115.1	Scaffold	2689.35	AWOL01	929607	0
Nicotiana paniculata	Eukaryota; Plants; Land Plants	GCA_005239505.1	Scaffold	2190.56	MDKI01	181977	0
Nicotiana rustica	Eukaryota; Plants; Land Plants	GCA_005239535.1	Scaffold	4231.29	MDKG01	337581	0
Nicotiana sylvestris	Eukaryota; Plants; Land Plants	GCA_000393655.1	Scaffold	2221.99	ASAF01	253918	48160
Nicotiana tabacum	Eukaryota; Plants; Land Plants	GCA_000715135.1	Scaffold	3643.47	AYMY01	168247	84255
Nicotiana tabacum	Eukaryota; Plants; Land Plants	GCA_000715075.1	Scaffold	3732.64	AWOJ01	582565	0
Nicotiana tabacum	Eukaryota; Plants; Land Plants	GCA_000715095.1	Scaffold	3735.82	AWOK01	643545	0
Nicotiana tabacum	Eukaryota; Plants; Land Plants	GCA_002210045.1	Scaffold	4646.65	NCAA01	937112	0
Nicotiana tomentosiformis	Eukaryota; Plants; Land Plants	GCA_000390325.2	Scaffold	1688.47	ASAG01	159548	48963
Nicotiana undulata	Eukaryota; Plants; Land Plants	GCA_005239495.1	Scaffold	1914.3	MDKH01	117566	0
Nissolia schottii	Eukaryota; Plants; Land Plants	GCA_003254905.1	Scaffold	466.099	QANU01	116213	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Noccaea caerulescens	Eukaryota; Plants; Land Plants	GCA_900406465.1	Scaffold	140.792	OVBC01	19808	0
Noccaea goesingensis	Eukaryota; Plants; Land Plants	GCA_900406475.1	Scaffold	150.323	OVBG01	183017	0
Nothapodytes nimmoniana	Eukaryota; Plants; Land Plants	GCA_002091855.1	Contig	1.36527	BDGC01	2301	0
Nymphaea colorata	Eukaryota; Plants; Land Plants	GCA_902499525.1	Scaffold	409.931	CABVML01	806	0
Nymphaea colorata	Eukaryota; Plants; Land Plants	GCA_008831285.1	Chromosome	408.397	VYXN01	799	0
Nyssa sinensis	Eukaryota; Plants; Land Plants	GCA_008638375.1	Chromosome	1001.45	VIRR01	654	36241
Ochetophila trinervis	Eukaryota; Plants; Land Plants	GCA_003254975.1	Scaffold	309.116	QANX01	8237	0
Ocimum tenuiflorum	Eukaryota; Plants; Land Plants	GCA_001278415.1	Contig	332.617	AYJT01	121993	0
Ocimum tenuiflorum	Eukaryota; Plants; Land Plants	GCA_001748785.1	Contig	311.125	JQCZ01	230018	0
Odon tarrhena argentea	Eukaryota; Plants; Land Plants	GCA_900406245.1	Scaffold	183.186	OVAF01	32097	0
Oenanthe javanica	Eukaryota; Plants; Land Plants	GCA_008931105.1	Scaffold	1278.51	QRFB01	149923	0
Olea europaea subsp. europaea	Eukaryota; Plants; Land Plants	GCA_900603015.1	Scaffold	1318.65	UWJE01	11038	0
Olea europaea var. sylvestris	Eukaryota; Plants; Land Plants	GCA_002742605.1	Chromosome	1141.15	MSRW01	41226	58334
Oropetium thomaeum	Eukaryota; Plants; Land Plants	GCA_001182835.1	Contig	243.175	LFJQ01	625	0

Oryza barthii	Eukaryota; Plants; Land Plants	GCA_002926215.1	Scaffold	295.586	PQXR01	46663	0
Oryza barthii	Eukaryota; Plants; Land Plants	GCA_002926235.1	Scaffold	294.191	PQXQ01	52511	0
Oryza barthii	Eukaryota; Plants; Land Plants	GCA_003020155.1	Scaffold	292.235	PTLQ01	66245	0
Oryza barthii	Eukaryota; Plants; Land Plants	GCA_000182155.3	Chromosome	308.272	ABRL02	12	0
Oryza brachyantha	Eukaryota; Plants; Land Plants	GCA_000710545.1	Chromosome	14.4404	JNWF01	1	0
Oryza brachyantha	Eukaryota; Plants; Land Plants	GCA_000231095.2	Chromosome	259.908	AGAT01	2491	26803
Oryza glaberrima	Eukaryota; Plants; Land Plants	GCA_000147395.2	Scaffold	303.295	ADWL01	25599	0
Oryza glumipatula	Eukaryota; Plants; Land Plants	GCA_000576495.1	Chromosome	372.86	ALNU02	12	0
Oryza longistaminata	Eukaryota; Plants; Land Plants	GCA_000789195.1	Scaffold	326.443	AMDW01	60198	0
Oryza longistaminata	Eukaryota; Plants; Land Plants	GCA_001514335.2	Chromosome	362.064	LQBC01	11745	0
Oryza meridionalis	Eukaryota; Plants; Land Plants	GCA_001551795.1	Contig	354.611	LONC01	3249	0
Oryza meridionalis	Eukaryota; Plants; Land Plants	GCA_000338895.2	Chromosome	335.668	ALNW02	12	0
Oryza meyeriana var. granulata	Eukaryota; Plants; Land Plants	GCA_005223365.1	Scaffold	736.649	SPHZ01	2389	0
Oryza meyeriana var. granulata	Eukaryota; Plants; Land Plants	GCA_003991445.1	Contig	776.957	RYFJ01	4618	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Oryza meyeriana var. granulata	Eukaryota; Plants; Land Plants	GCA_000325645.2	Scaffold	35.2457	ALNT01	1	0
Oryza minuta	Eukaryota; Plants; Land Plants	GCA_000632695.1	Chromosome	45.1659	JJNN01	2	0
Oryza officinalis	Eukaryota; Plants; Land Plants	GCA_008326285.1	Scaffold	584.134	BDMV01	91	0
Oryza officinalis	Eukaryota; Plants; Land Plants	GCA_000717455.1	Chromosome	26.1885	JJMQ01	1	0
Oryza punctata	Eukaryota; Plants; Land Plants	GCA_000710525.1	Chromosome	22.4654	JNWE01	1	0
Oryza punctata	Eukaryota; Plants; Land Plants	GCA_000573905.1	Chromosome	393.817	AVCL01	12	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_000817225.1	Scaffold	339.177	CBQP01	3818	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_001551805.1	Contig	384.518	LONB01	2582	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609075.1	Contig	0.391959	UXAX01	64	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609175.1	Contig	0.439665	UXBI01	116	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609365.1	Contig	0.435203	UXBZ01	67	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609345.1	Contig	0.430311	UXBX01	112	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609235.1	Contig	0.431171	UXBN01	130	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609435.1	Contig	0.460371	UXCI01	141	0

Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609385.1	Contig	0.522775	UXCD01	399	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609605.1	Contig	0.479941	UXCZ01	272	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609155.1	Contig	0.451183	UXBF01	185	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609195.1	Contig	0.444689	UXBK01	168	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609375.1	Contig	0.418576	UXCA01	72	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609635.1	Contig	0.437219	UXDC01	115	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609335.1	Contig	0.426264	UXBY01	80	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609565.1	Contig	0.416343	UXCV01	99	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609455.1	Contig	0.437726	UXCL01	172	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609145.1	Contig	0.4324	UXBJ01	128	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609205.1	Contig	0.429589	UXBM01	129	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609535.1	Contig	0.425584	UXCS01	176	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609445.1	Contig	0.428734	UXCJ01	87	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609595.1	Contig	0.445408	UXDA01	162	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609645.1	Contig	0.448673	UXDD01	155	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609285.1	Contig	0.420471	UXBV01	143	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609165.1	Contig	0.426734	UXBG01	121	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609325.1	Contig	0.456272	UXCH01	231	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609355.1	Contig	0.427134	UXCB01	114	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609415.1	Contig	0.439065	UXCE01	125	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609575.1	Contig	0.426419	UXCW01	113	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609125.1	Contig	0.461807	UXBC01	223	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609475.1	Contig	0.435078	UXCN01	159	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609215.1	Contig	0.433062	UXBO01	157	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609395.1	Contig	0.399871	UXCF01	66	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609585.1	Contig	0.433896	UXCY01	137	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609555.1	Contig	0.423266	UXCT01	170	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609295.1	Contig	0.403327	UXBR01	109	0

Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609545.1	Contig	0.436432	UXCU01	221	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609185.1	Contig	0.434658	UXBH01	155	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609525.1	Contig	0.44213	UXCQ01	233	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609105.1	Contig	0.42703	UXBD01	130	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609615.1	Contig	0.4457	UXCX01	188	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609505.1	Contig	0.49038	UXCR01	321	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609135.1	Contig	0.424383	UXBE01	108	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609495.1	Contig	0.429759	UXCM01	130	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609305.1	Contig	0.531837	UXBS01	488	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609625.1	Contig	0.442979	UXDB01	201	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609265.1	Contig	0.430376	UXBQ01	146	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609245.1	Contig	0.441818	UXBP01	94	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609405.1	Contig	0.466007	UXCC01	262	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609225.1	Contig	0.428872	UXBL01	138	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609465.1	Contig	0.420063	UXCK01	121	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609255.1	Contig	0.551867	UXBT01	457	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609515.1	Contig	0.438855	UXCP01	218	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609085.1	Contig	0.467544	UXAY01	265	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609275.1	Contig	0.552388	UXBU01	559	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609095.1	Contig	0.4166	UXBA01	216	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609315.1	Contig	0.576561	UXBW01	564	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609485.1	Contig	0.512907	UXCO01	440	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609425.1	Contig	0.5568	UXCG01	519	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609115.1	Contig	0.416778	UXBB01	188	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_900609665.1	Contig	0.521122	UXDG01	471	0
Oryza rufipogon	Eukaryota; Plants; Land Plants	GCA_000700045.1	Chromosome	12.7409	JNHC01	1	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_001648735.1	Scaffold	307.225	LVCG01	55637	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_001648745.1	Scaffold	295.39	LVCH01	64800	0

Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609875.1	Contig	0.418428	UXEA01	94	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609755.1	Contig	0.412758	UXDO01	77	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900610885.1	Contig	0.415805	UXHM01	101	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609945.1	Contig	0.420498	UXEH01	97	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609745.1	Contig	0.412413	UXDL01	77	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609855.1	Contig	0.424341	UXDZ01	111	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609765.1	Contig	0.426436	UXDS01	105	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609895.1	Contig	0.420731	UXED01	102	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609705.1	Contig	0.413182	UXDI01	83	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609835.1	Contig	0.414506	UXDX01	86	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609655.1	Contig	0.4229	UXDE01	95	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609715.1	Contig	0.421534	UXDK01	109	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609735.1	Contig	0.412833	UXDM01	71	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609915.1	Contig	0.417059	UXEE01	85	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609955.1	Contig	0.41464	UXEI01	83	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609885.1	Contig	0.415361	UXEC01	81	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900610895.1	Contig	0.414252	UXHO01	91	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609905.1	Contig	0.416009	UXEB01	104	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609865.1	Contig	0.4183	UXDY01	102	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609845.1	Contig	0.424074	UXDW01	114	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609685.1	Contig	0.415611	UXDH01	101	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609925.1	Contig	0.412137	UXEF01	98	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609695.1	Contig	0.414261	UXDJ01	76	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609775.1	Contig	0.411428	UXDP01	78	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609825.1	Contig	0.415893	UXDV01	83	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609725.1	Contig	0.413692	UXDN01	95	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609805.1	Contig	0.411297	UXDU01	88	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609675.1	Contig	0.413512	UXDF01	83	0

Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609935.1	Contig	0.415952	UXEG01	100	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609815.1	Contig	0.405341	UXDT01	71	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609785.1	Contig	0.42347	UXDQ01	148	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_900609795.1	Contig	0.422231	UXDR01	125	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_003865215.1	Chromosome	395.354	PKRX01	127	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_004007595.1	Chromosome	377.604	RPSM01	615	0
Oryza sativa	Eukaryota; Plants; Land Plants	GCA_004348155.2	Chromosome	415.393	QQAJ01	367	0
Oryza sativa aus subgroup	Eukaryota; Plants; Land Plants	GCA_001952365.2	Chromosome	372.203	LWDA01	1312	0
Oryza sativa f. spontanea	Eukaryota; Plants; Land Plants	GCA_006942195.1	Scaffold	377.675	QKSA01	350	0
Oryza sativa f. spontanea	Eukaryota; Plants; Land Plants	GCA_000710535.2	Chromosome	19.4244	JNWG02	1	0
Oryza sativa f. spontanea	Eukaryota; Plants; Land Plants	GCA_000576065.1	Chromosome	337.95	AWHD01	12	0
Oryza sativa Indica Group	Eukaryota; Plants; Land Plants	GCA_001623365.2	Chromosome	387.424	LNNK02	19	0
Oryza sativa Indica Group	Eukaryota; Plants; Land Plants	GCA_001623345.2	Chromosome	387.326	LNNJ02	20	0
Oryza sativa Indica Group	Eukaryota; Plants; Land Plants	GCA_001889745.1	Chromosome	389.088	MPPV01	66	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Oryza sativa Indica Group	Eukaryota; Plants; Land Plants	GCA_001618795.1	Chromosome	386.486	LBBA01	8481	0
Oryza sativa Indica Group	Eukaryota; Plants; Land Plants	GCA_000725085.2	Chromosome	389.753	AZTA02	15907	0
Oryza sativa Indica Group	Eukaryota; Plants; Land Plants	GCA_001618785.1	Chromosome	398.762	LBAZ01	11486	0
Oryza sativa Indica Group	Eukaryota; Plants; Land Plants	GCA_000004655.2	Chromosome	426.337	AAAA02	10627	37358
Oryza sativa Indica Group	Eukaryota; Plants; Land Plants	GCA_003449045.1	Contig	388.772	QWGD01	410	0
Oryza sativa Indica Group	Eukaryota; Plants; Land Plants	GCA_001611195.1	Scaffold	351.225	LQHG01	39669	0
Oryza sativa Indica Group	Eukaryota; Plants; Land Plants	GCA_001611255.1	Scaffold	331.819	LQHF01	75421	0
Oryza sativa Indica Group	Eukaryota; Plants; Land Plants	GCA_001611235.1	Scaffold	352.227	LQHE01	116212	0
Oryza sativa Indica Group	Eukaryota; Plants; Land Plants	GCA_006992885.1	Scaffold	281.326	SWLY01	91264	0
Oryza sativa Japonica Group	Eukaryota; Plants; Land Plants	GCA_003449065.1	Contig	378.097	QWGC01	144	0
Oryza sativa Japonica Group	Eukaryota; Plants; Land Plants	GCA_002573525.1	Contig	418.901	PDFQ01	588	0
Oryza sativa Japonica Group	Eukaryota; Plants; Land Plants	GCA_003865235.1	Chromosome	379.626	PKRW01	115	0
Oryza sativa Japonica Group	Eukaryota; Plants; Land Plants	GCA_000817635.1	Chromosome	337.74	JSUG01	1739	0
Oryza sativa Japonica Group	Eukaryota; Plants; Land Plants	GCA_000817615.1	Chromosome	342.028	JSUF01	1160	0

Oryza sativa Japonica Group	Eukaryota; Plants; Land Plants	GCA_000149285.1	Chromosome	391.148	AACV01	7777	35394
Oryza sativa Japonica Group	Eukaryota; Plants; Land Plants	GCA_000321445.1	Chromosome	382.627	BACJ01	12	0
Oryza sativa Japonica Group	Eukaryota; Plants; Land Plants	GCA_000164945.1	Chromosome	382.151	BABO01	12	0
Pachycereus pringlei	Eukaryota; Plants; Land Plants	GCA_002740445.1	Scaffold	629.656	NCQS01	171584	0
Panicum hallii	Eukaryota; Plants; Land Plants	GCA_002211085.2	Chromosome	535.889	NCQW02	1027	37612
Panicum hallii var. hallii	Eukaryota; Plants; Land Plants	GCA_003061485.1	Chromosome	487.474	QAVV01	144	42523
Panicum miliaceum	Eukaryota; Plants; Land Plants	GCA_002895445.2	Chromosome	848.352	PPDP02	466	0
Panicum miliaceum	Eukaryota; Plants; Land Plants	GCA_003046395.2	Chromosome	854.793	PQIB02	1306	55964
Papaver somniferum	Eukaryota; Plants; Land Plants	GCA_003573695.1	Chromosome	2715.53	PUWZ01	34381	84179
Parasponia andersonii	Eukaryota; Plants; Land Plants	GCA_002914805.1	Scaffold	475.834	JXTB01	2732	37227
Passiflora edulis	Eukaryota; Plants; Land Plants	GCA_002156105.1	Scaffold	165.657	MUZT01	234012	0
Penstemon barbatus	Eukaryota; Plants; Land Plants	GCA_003313485.1	Contig	696.306	QOIQ01	18827	0
Penstemon centranthifolius	Eukaryota; Plants; Land Plants	GCA_000737435.1	Contig	4.47159	JPFH01	6761	0
Penstemon cyananthus	Eukaryota; Plants; Land Plants	GCA_000281005.1	Contig	4.62226	AKKG01	9712	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Penstemon davidsonii	Eukaryota; Plants; Land Plants	GCA_000280985.1	Contig	2.37523	AKKI01	4880	0
Penstemon dissectus	Eukaryota; Plants; Land Plants	GCA_000280965.1	Contig	2.62809	AKKH01	5361	0
Penstemon fruticosus	Eukaryota; Plants; Land Plants	GCA_000281025.1	Contig	2.31904	AKKJ01	4770	0
Penstemon grinnellii	Eukaryota; Plants; Land Plants	GCA_000737425.1	Contig	3.66352	JPFI01	5523	0
Pereskia humboldtii	Eukaryota; Plants; Land Plants	GCA_002740485.1	Scaffold	414.047	NCQU01	126352	0
Perilla citriodora	Eukaryota; Plants; Land Plants	GCA_004303085.1	Scaffold	618.797	SDAM01	29924	0
Persea americana	Eukaryota; Plants; Land Plants	GCA_008087245.1	Contig	912.698	SDSS01	8135	0
Persea americana	Eukaryota; Plants; Land Plants	GCA_002908915.1	Contig	446.756	NXHZ01	5000	0
Persea americana var. drymifolia	Eukaryota; Plants; Land Plants	GCA_008033785.1	Scaffold	820.369	SDXN01	43777	0
Phalaenopsis aphrodite	Eukaryota; Plants; Land Plants	GCA_003013225.1	Scaffold	1025.1	NEWO01	13732	0
Phalaenopsis equestris	Eukaryota; Plants; Land Plants	GCA_001263595.1	Scaffold	1064.2	APLD01	89584	29894
Phalaenopsis hybrid cultivar	Eukaryota; Plants; Land Plants	GCA_002079205.1	Scaffold	2687.66	JXCR01	149149	0
Phaseolus coccineus subsp. coccineus	Eukaryota; Plants; Land Plants	GCA_003122825.1	Scaffold	371.086	QBDZ01	192921	0
Phaseolus vulgaris	Eukaryota; Plants; Land Plants	GCA_001517995.1	Chromosome	549.748	LPQZ01	68335	0

Phaseolus vulgaris	Eukaryota; Plants; Land Plants	GCA_000499845.1	Chromosome	521.077	ANNZ01	708	32720
Phoenix dactylifera	Eukaryota; Plants; Land Plants	GCA_000413155.1	Scaffold	556.481	ATBV01	80317	40634
Phoenix dactylifera	Eukaryota; Plants; Land Plants	GCA_000181215.2	Scaffold	381.563	ACYX02	57277	0
Phoenix dactylifera	Eukaryota; Plants; Land Plants	GCA_007821505.1	Scaffold	454.367	PEFZ01	252335	0
Physaria acutifolia	Eukaryota; Plants; Land Plants	GCA_900406485.1	Scaffold	199.442	OVBH01	276837	0
Physaria fendleri	Eukaryota; Plants; Land Plants	GCA_900406525.1	Scaffold	331.342	OVBV01	99932	0
Physaria ovalifolia	Eukaryota; Plants; Land Plants	GCA_900406505.1	Scaffold	290.267	OVBY01	335399	0
Physcomitrella patens	Eukaryota; Plants; Land Plants	GCA_000002425.2	Chromosome	472.081	ABEU02	359	48022
Picea abies	Eukaryota; Plants; Land Plants	GCA_900067695.1	Scaffold	11961.4	CBVK01	11340369	0
Picea abies var. abies	Eukaryota; Plants; Land Plants	GCA_900491625.1	Scaffold	42.7831	UETF01	41150	0
Picea glauca	Eukaryota; Plants; Land Plants	GCA_000411955.5	Scaffold	24633.1	ALWZ04	3033322	6445
Picea glauca	Eukaryota; Plants; Land Plants	GCA_000966675.1	Scaffold	26936.2	JZKD01	3353683	0
Picea glauca	Eukaryota; Plants; Land Plants	GCA_001687225.1	Contig	258.272	LDPM01	222034	0
Pinus lambertiana	Eukaryota; Plants; Land Plants	GCA_001447015.2	Scaffold	27602.7	LMTP01	4253097	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Pinus sylvestris	Eukaryota; Plants; Land Plants	GCA_900143225.1	Contig	0.985624	FRDG01	224	0
Pinus taeda	Eukaryota; Plants; Land Plants	GCA_000404065.3	Scaffold	22103.6	APFE03	1760464	0
Pistacia vera	Eukaryota; Plants; Land Plants	GCA_008641045.1	Scaffold	671.28	VTWI01	1865	82
Pisum sativum	Eukaryota; Plants; Land Plants	GCA_003013575.1	Scaffold	4275.93	PUCA01	5449423	0
Platycodon grandiflorus	Eukaryota; Plants; Land Plants	GCA_004681165.1	Scaffold	680.178	SPEA01	4816	0
Pleurozium schreberi	Eukaryota; Plants; Land Plants	GCA_006891605.1	Contig	220.032	VACF01	2689	0
Pogostemon cablin	Eukaryota; Plants; Land Plants	GCA_003675935.1	Scaffold	1916.69	QKXD01	41698	0
Populus alba	Eukaryota; Plants; Land Plants	GCA_005239225.1	Contig	416.961	RCHU01	1287	32959
Populus euphratica	Eukaryota; Plants; Land Plants	GCA_000495115.1	Scaffold	496.033	AOFL01	9615	49760
Populus simonii	Eukaryota; Plants; Land Plants	GCA_007827005.2	Chromosome	441.407	VJNQ02	686	0
Populus trichocarpa	Eukaryota; Plants; Land Plants	GCA_000002775.3	Chromosome	434.29	AARH03	1694	51717
Primula veris	Eukaryota; Plants; Land Plants	GCA_000788445.1	Scaffold	309.693	JTKG01	8756	0
Primula vulgaris	Eukaryota; Plants; Land Plants	GCA_001077355.1	Scaffold	1.50478	CDJJ02	229	0
Primula vulgaris	Eukaryota; Plants; Land Plants	GCA_001403715.1	Scaffold	1.50478	CYSU01	229	0

Prosopis alba	Eukaryota; Plants; Land Plants	GCA_004799145.1	Contig	707.162	SMJV01	6087	57572
Prunus avium	Eukaryota; Plants; Land Plants	GCA_002207925.1	Scaffold	272.362	BDGV01	10148	35009
Prunus avium	Eukaryota; Plants; Land Plants	GCA_003946875.1	Contig	287.192	QXJJ01	1540	0
Prunus dulcis	Eukaryota; Plants; Land Plants	GCA_902201215.1	Chromosome	227.599	CABIKO01	691	32556
Prunus mume	Eukaryota; Plants; Land Plants	GCA_000346735.1	Chromosome	234.03	AOHF01	8626	29705
Prunus persica	Eukaryota; Plants; Land Plants	GCA_000218175.1	Scaffold	214.225	AEJG01	30834	0
Prunus persica	Eukaryota; Plants; Land Plants	GCA_000218215.1	Scaffold	207.185	AEKV01	43890	0
Prunus persica	Eukaryota; Plants; Land Plants	GCA_000218195.1	Scaffold	211.308	AEKW01	35219	0
Prunus persica	Eukaryota; Plants; Land Plants	GCA_000346465.2	Chromosome	227.569	AKXU02	192	32595
Prunus yedoensis	Eukaryota; Plants; Land Plants	GCA_005406145.1	Contig	690.106	BJCG01	4571	0
Prunus yedoensis var. nudiflora	Eukaryota; Plants; Land Plants	GCA_002966975.2	Scaffold	319.21	PJQY01	4016	41294
Prunus yedoensis var. nudiflora	Eukaryota; Plants; Land Plants	GCA_900382725.1	Scaffold	319.21	OSDV01	4016	0
Pseudotsuga menziesii	Eukaryota; Plants; Land Plants	GCA_001517045.1	Scaffold	14673.2	LPNX01	1236665	0
Pseudoturritis turrita	Eukaryota; Plants; Land Plants	GCA_900406555.1	Contig	321.563	OVBL01	1111	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Pseudoturritis turrita	Eukaryota; Plants; Land Plants	GCA_900406515.1	Scaffold	263.513	OVBK01	25008	0
Psidium guajava	Eukaryota; Plants; Land Plants	GCA_002914565.1	Contig	386.852	NTGF01	4728	0
Pterocarya stenoptera	Eukaryota; Plants; Land Plants	GCA_003123785.1	Scaffold	955.601	QEOT01	124315	0
Punica granatum	Eukaryota; Plants; Land Plants	GCA_002864125.1	Scaffold	274.043	MTJX01	2117	0
Punica granatum	Eukaryota; Plants; Land Plants	GCA_002201585.1	Scaffold	296.383	MTKT01	17405	29127
Punica granatum	Eukaryota; Plants; Land Plants	GCA_002837095.1	Scaffold	380.178	PGOL01	45308	50476
Punica granatum	Eukaryota; Plants; Land Plants	GCA_007655135.2	Chromosome	320.336	MABG02	473	0
Purshia tridentata	Eukaryota; Plants; Land Plants	GCA_003254885.1	Scaffold	175.971	QANT01	9353	0
Pyrus betulifolia	Eukaryota; Plants; Land Plants	GCA_007844245.1	Chromosome	532.747	VDML01	139	0
Pyrus x bretschneideri	Eukaryota; Plants; Land Plants	GCA_000315295.1	Scaffold	508.551	AJSU01	2182	47086
Quercus lobata	Eukaryota; Plants; Land Plants	GCA_001633185.2	Chromosome	846.07	LRBV02	2010	53228
Quercus robur	Eukaryota; Plants; Land Plants	GCA_900291515.1	Scaffold	814.336	OLKR01	550	0
Quercus robur	Eukaryota; Plants; Land Plants	GCA_003013145.1	Scaffold	719.602	PVWZ01	84416	0
Quercus suber	Eukaryota; Plants; Land Plants	GCA_002906115.1	Scaffold	953.299	PKMF01	23344	59614

Quillaja saponaria	Eukaryota; Plants; Land Plants	GCA_003338715.1	Contig	248.908	PVLG01	48349	0
Raddia distichophylla	Eukaryota; Plants; Land Plants	GCA_005191435.1	Scaffold	580.833	SPJY01	38257	0
Raparia bulbosa	Eukaryota; Plants; Land Plants	GCA_900406535.1	Scaffold	152.784	OVBJ01	13874	0
Raphanus raphanistrum subsp. raphanistrum	Eukaryota; Plants; Land Plants	GCA_000769845.1	Contig	253.834	JRQH01	64732	0
Raphanus sativus	Eukaryota; Plants; Land Plants	GCA_000801105.2	Scaffold	426.614	JRUI02	10676	61216
Raphanus sativus	Eukaryota; Plants; Land Plants	GCA_000715565.1	Scaffold	402.328	BAUK01	76592	0
Raphanus sativus	Eukaryota; Plants; Land Plants	GCA_001047155.1	Scaffold	383.105	BAOO01	40123	0
Raphanus sativus	Eukaryota; Plants; Land Plants	GCA_002197605.1	Chromosome	382.79	JSDR01	44239	0
Rhamnella rubrinervis	Eukaryota; Plants; Land Plants	GCA_007844105.1	Chromosome	245.336	VOIH01	133	0
Rhazya stricta	Eukaryota; Plants; Land Plants	GCA_001752375.1	Scaffold	274.354	MEJB01	979	0
Rhizophora apiculata	Eukaryota; Plants; Land Plants	GCA_900174605.1	Scaffold	232.055	FWPW01	142	0
Rhizophora apiculata	Eukaryota; Plants; Land Plants	GCA_900004065.1	Scaffold	232.431	CELW01	45996	0
Rhodamnia argentea	Eukaryota; Plants; Land Plants	GCA_900635035.1	Scaffold	414.816	CAAAGQ01	15781	42570
Rhodoleia championii	Eukaryota; Plants; Land Plants	GCA_008932045.1	Contig	105.726	VMOD01	278658	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Ricinus communis	Eukaryota; Plants; Land Plants	GCA_000151685.2	Scaffold	350.622	AASG02	25763	28584
Rosa chinensis	Eukaryota; Plants; Land Plants	GCA_002994745.1	Chromosome	513.854	PDCK01	45	45097
Rosa lucieae	Eukaryota; Plants; Land Plants	GCA_006954505.1	Scaffold	786.105	RQIQ01	500476	0
Rosa multiflora	Eukaryota; Plants; Land Plants	GCA_002564525.1	Scaffold	739.638	BDJD01	83189	0
Rosa x damascena	Eukaryota; Plants; Land Plants	GCA_001662545.1	Scaffold	711.72	LYNE01	307872	0
Ruellia speciosa	Eukaryota; Plants; Land Plants	GCA_001909325.1	Contig	740.036	MAYD01	794288	0
Saccharum hybrid cultivar	Eukaryota; Plants; Land Plants	GCA_900465005.1	Scaffold	530.66	UBIK01	5708	0
Saccharum hybrid cultivar SP80-3280	Eukaryota; Plants; Land Plants	GCA_008692665.1	Scaffold	4014.93	QPEU01	398353	0
Saccharum hybrid cultivar SP80-3280	Eukaryota; Plants; Land Plants	GCA_002018215.1	Contig	1169.95	JXQF01	199028	0
Saccharum hybrid cultivar SP80-3280	Eukaryota; Plants; Land Plants	GCA_009173535.1	Scaffold	49.3852	PYBL01	461	0
Saccharum spontaneum	Eukaryota; Plants; Land Plants	GCA_900500655.1	Contig	3924.19	UINE01	75981	0
Saccharum spontaneum	Eukaryota; Plants; Land Plants	GCA_003544955.1	Chromosome	3133.29	QVOL01	15303	0
Salix brachista	Eukaryota; Plants; Land Plants	GCA_009078335.1	Chromosome	339.588	VDCV01	30	30209
Salvia splendens	Eukaryota; Plants; Land Plants	GCA_004379255.1	Scaffold	809.16	PNBA01	1525	53354

Santalum album	Eukaryota; Plants; Land Plants	GCA_002911635.1	Contig	196.101	NXEK01	180	0
Santalum album	Eukaryota; Plants; Land Plants	GCA_002925775.1	Scaffold	220.961	LOCJ01	12821	0
Schrenkiella parvula	Eukaryota; Plants; Land Plants	GCA_000218505.1	Chromosome	137.073	AFAN01	1463	0
Scutellaria baicalensis	Eukaryota; Plants; Land Plants	GCA_005771605.1	Chromosome	386.674	VALJ01	114	0
Secale cereale	Eukaryota; Plants; Land Plants	GCA_900079665.1	Scaffold	1684.93	FKKI01	1581707	0
Secale cereale	Eukaryota; Plants; Land Plants	GCA_900002355.1	Scaffold	1684.93	CCJQ01	1581707	0
Sedum album	Eukaryota; Plants; Land Plants	GCA_006409495.1	Contig	302.251	QZGG01	6038	0
Selaginella kraussiana	Eukaryota; Plants; Land Plants	GCA_001021135.1	Scaffold	114.503	LDJE01	105914	0
Selaginella moellendorffii	Eukaryota; Plants; Land Plants	GCA_000143415.2	Scaffold	212.315	ADFJ01	757	45247
Selaginella tamariscina	Eukaryota; Plants; Land Plants	GCA_003024785.1	Scaffold	300.729	PUQB01	1391	0
Sequoia sempervirens	Eukaryota; Plants; Land Plants	GCA_007258455.1	Scaffold	26537.2	VDFB01	517852	0
Sequoia dendrongiganteum	Eukaryota; Plants; Land Plants	GCA_007115665.1	Scaffold	8122.13	VCHN01	39798	0
Sesamum indicum	Eukaryota; Plants; Land Plants	GCA_001692995.1	Scaffold	210.758	MBSK01	5868	0
Sesamum indicum	Eukaryota; Plants; Land Plants	GCA_003268515.1	Scaffold	242.679	LUAT01	48805	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Sesamum indicum	Eukaryota; Plants; Land Plants	GCA_000975565.1	Scaffold	340.464	JPLX01	76023	0
Sesamum indicum	Eukaryota; Plants; Land Plants	GCA_000512975.1	Chromosome	275.059	APMJ01	16369	35410
Setaria italica	Eukaryota; Plants; Land Plants	GCA_001652605.1	Chromosome	477.542	LWRS01	2689	0
Setaria italica	Eukaryota; Plants; Land Plants	GCA_000263155.2	Chromosome	405.868	AGNK02	337	35844
Setaria viridis	Eukaryota; Plants; Land Plants	GCA_005286985.1	Chromosome	395.732	SNSE01	75	52459
Silene latifolia	Eukaryota; Plants; Land Plants	GCA_003260165.1	Scaffold	1185.09	QBIE01	319506	0
Silene latifolia	Eukaryota; Plants; Land Plants	GCA_900095335.1	Contig	36.0486	FMHP01	46178	0
Silene latifolia subsp. alba	Eukaryota; Plants; Land Plants	GCA_001412135.1	Scaffold	665.279	LHUT01	307720	0
Silphium perfoliatum	Eukaryota; Plants; Land Plants	GCA_900538075.1	Contig	121.712	UXAI01	1197534	0
Silybum marianum	Eukaryota; Plants; Land Plants	GCA_001541825.1	Contig	1477.57	LMWD01	258575	0
Sisymbrium altissimum	Eukaryota; Plants; Land Plants	GCA_900406495.1	Scaffold	178.647	OVBI01	14597	0
Sisymbrium irio	Eukaryota; Plants; Land Plants	GCA_000411075.1	Scaffold	245.55	ASZH01	21357	0
Solanum americanum	Eukaryota; Plants; Land Plants	GCA_900188915.1	Contig	9.01369	FYFB01	837	0
Solanum americanum	Eukaryota; Plants; Land Plants	GCA_900188785.1	Contig	7.74921	FYHF01	1085	0

Solanum americanum	Eukaryota; Plants; Land Plants	GCA_900188895.1	Contig	7.74473	FYHB01	1085	0
Solanum americanum	Eukaryota; Plants; Land Plants	GCA_900188885.1	Contig	7.6066	FYHD01	1085	0
Solanum americanum	Eukaryota; Plants; Land Plants	GCA_900198685.1	Contig	8.308	FZPQ01	1460	0
Solanum americanum	Eukaryota; Plants; Land Plants	GCA_900188835.1	Contig	9.83576	FYHH01	1483	0
Solanum arcanum	Eukaryota; Plants; Land Plants	GCA_000612985.1	Contig	665.187	CBYQ01	46594	0
Solanum chilense	Eukaryota; Plants; Land Plants	GCA_006013705.1	Scaffold	913.881	RXGB01	81304	19
Solanum commersonii	Eukaryota; Plants; Land Plants	GCA_001239805.1	Scaffold	729.603	JXZD01	63664	0
Solanum habrochaites	Eukaryota; Plants; Land Plants	GCA_000577655.1	Contig	724.285	CBYS01	42990	0
Solanum lycopersicum	Eukaryota; Plants; Land Plants	GCA_000181095.1	Scaffold	540.589	BABP01	100783	0
Solanum lycopersicum	Eukaryota; Plants; Land Plants	GCA_000325825.1	Scaffold	0.575198	AFYB01	195	0
Solanum lycopersicum	Eukaryota; Plants; Land Plants	GCA_000188115.3	Chromosome	828.349	AEKE03	3150	37660
Solanum melongena	Eukaryota; Plants; Land Plants	GCA_000787875.1	Scaffold	833.081	BAUE01	33873	0
Solanum pennellii	Eukaryota; Plants; Land Plants	GCA_000577875.1	Contig	720.458	CBYR01	57205	0
Solanum pennellii	Eukaryota; Plants; Land Plants	GCA_000820945.1	Contig	720.458	CCXL01	57205	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Solanum pimpinellifolium	Eukaryota; Plants; Land Plants	GCA_003660305.1	Scaffold	748.694	NRDK01	48863	0
Solanum pimpinellifolium	Eukaryota; Plants; Land Plants	GCA_000230315.1	Contig	688.247	AGFK01	309180	0
Solanum tuberosum	Eukaryota; Plants; Land Plants	GCA_000226075.1	Scaffold	705.934	AEWC01	14854	37966
Solanum tuberosum	Eukaryota; Plants; Land Plants	GCA_900004685.1	Contig	90.4582	CVMJ01	39446	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185145.1	Scaffold	730.142	FYAA01	224100	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185165.1	Scaffold	667.086	FXZQ01	2343	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185285.1	Contig	659.291	FXZO01	2461	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185275.1	Scaffold	662.264	FXZP01	1331	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185325.1	Scaffold	659.429	FXZS01	1377	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185305.1	Scaffold	716.55	FXZW01	4814	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185245.1	Contig	715.934	FXZV01	5571	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185335.1	Scaffold	749.835	FXZU01	5164	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185175.1	Scaffold	764.063	FXZT01	7840	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185215.1	Contig	722.285	FXZR01	8138	0

Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185185.1	Scaffold	751.585	FYAB01	321725	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185195.1	Scaffold	740.946	FYAC01	170623	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185295.1	Scaffold	728.86	FYAD01	228374	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185265.1	Scaffold	729.31	FYAF01	224108	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185155.1	Scaffold	690.417	FXZY01	22476	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185205.1	Scaffold	688.737	FXZX01	22492	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185315.1	Scaffold	710.407	FYAG01	246082	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185225.1	Scaffold	730.903	FXZZ01	182474	0
Solanum verrucosum	Eukaryota; Plants; Land Plants	GCA_900185235.1	Scaffold	759.168	FYAE01	569872	0
Sorghum bicolor	Eukaryota; Plants; Land Plants	GCA_003482435.1	Scaffold	666.155	QWKM01	308	0
Sorghum bicolor	Eukaryota; Plants; Land Plants	GCA_008000285.1	Contig	374.252	VOIB01	2657	0
Sorghum bicolor	Eukaryota; Plants; Land Plants	GCA_000236765.2	Contig	0.015475	AHAO01	16	16
Sorghum bicolor	Eukaryota; Plants; Land Plants	GCA_000236725.2	Contig	0.018494	AHAQ01	20	22
Sorghum bicolor	Eukaryota; Plants; Land Plants	GCA_000236745.2	Contig	0.021299	AHAP01	35	35

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Sorghum bicolor	Eukaryota; Plants; Land Plants	GCA_000003195.3	Chromosome	709.345	ABXC03	869	39248
Spatholobus suberectus	Eukaryota; Plants; Land Plants	GCA_004329165.1	Chromosome	798.47	QUWT01	816	31106
Spinacia oleracea	Eukaryota; Plants; Land Plants	GCA_002007265.1	Scaffold	869.946	LZYP01	78263	32794
Spinacia oleracea	Eukaryota; Plants; Land Plants	GCA_000510995.2	Scaffold	493.772	AYZV02	103502	23522
Spirodela polyrhiza	Eukaryota; Plants; Land Plants	GCA_900492545.1	Scaffold	138.592	UIDA01	20	0
Spirodela polyrhiza	Eukaryota; Plants; Land Plants	GCA_008360905.1	Scaffold	138.536	SWLF01	134	0
Spirodela polyrhiza	Eukaryota; Plants; Land Plants	GCA_900536055.1	Scaffold	142.661	UNPA01	2585	0
Spirodela polyrhiza	Eukaryota; Plants; Land Plants	GCA_000504445.1	Contig	132.009	ATDW01	16051	0
Sporobolus alterniflorus	Eukaryota; Plants; Land Plants	GCA_008808055.1	Contig	365.573	VSTD01	52184	0
Stenocereus thurberi	Eukaryota; Plants; Land Plants	GCA_002740465.1	Scaffold	853.348	NCQT01	159477	0
Striga asiatica	Eukaryota; Plants; Land Plants	GCA_008636005.1	Scaffold	471.563	BKCP01	13846	33426
Syzygium oleosum	Eukaryota; Plants; Land Plants	GCA_900635055.1	Scaffold	431.291	CAAAGS01	19039	38158
Tarenaya hassleriana	Eukaryota; Plants; Land Plants	GCA_000463585.1	Scaffold	249.93	AOUI01	12249	41094
Theobroma cacao	Eukaryota; Plants; Land Plants	GCA_000403535.1	Chromosome	345.994	ALXC01	814	44186

Theobroma cacao	Eukaryota; Plants; Land Plants	GCA_000208745.2	Chromosome	324.88	FLSQ01	431	30854
Thlaspi arvense	Eukaryota; Plants; Land Plants	GCA_000956625.1	Scaffold	343.012	AZNP01	6768	0
Trema orientale	Eukaryota; Plants; Land Plants	GCA_002914845.1	Scaffold	387.958	JXTC01	2756	35849
Trichopus zeylanicus subsp. travancoricus	Eukaryota; Plants; Land Plants	GCA_005019695.1	Scaffold	713.407	RXID01	22601	0
Trifolium medium	Eukaryota; Plants; Land Plants	GCA_003490085.1	Scaffold	492.653	LXQA01	1471389	0
Trifolium pratense	Eukaryota; Plants; Land Plants	GCA_900292005.1	Chromosome	351.622	OMTE01	38479	0
Trifolium pratense	Eukaryota; Plants; Land Plants	GCA_900079335.1	Chromosome	345.991	FKJA01	39051	0
Trifolium pratense	Eukaryota; Plants; Land Plants	GCA_000583005.2	Contig	304.972	ASHM01	267372	63850
Trifolium subterraneum	Eukaryota; Plants; Land Plants	GCA_001742945.1	Scaffold	471.834	BCLP01	27424	42059
Trifolium subterraneum	Eukaryota; Plants; Land Plants	GCA_002003065.1	Contig	392.71	BBPR01	968279	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_002220415.2	Contig	15344.7	NMPL02	279439	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_900241085.1	Scaffold	13916.9	OETA01	519179	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_900067645.1	Scaffold	13427.4	FAOM01	735943	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_900000045.1	Scaffold	9134.02	CCYC01	6870110	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_900067735.1	Scaffold	10058.1	CBTL01	11673940	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_001077335.1	Scaffold	58.5022	CBUC01	1450	250
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_002158495.1	Scaffold	567.21	MOLT01	10339	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_000818885.1	Scaffold	65.1111	JROL01	15144	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_002999095.1	Scaffold	574.295	PKRY01	9044	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_900236235.1	Scaffold	452.948	OEIT01	23433	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_002780475.1	Contig	240.199	NTGG01	127921	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_002780545.1	Contig	609.488	NTGH01	523543	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_002780565.1	Contig	564.3	NTGI01	565551	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_900235945.1	Scaffold	839.076	OEIJ01	767884	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_900235935.1	Scaffold	804.001	ODGO01	749802	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_001485685.1	Contig	44.4015	FAOV01	50000	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_000334095.1	Contig	3800.33	CALP01	5321847	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_000188135.1	Contig	159.087	AEOM01	311945	0

Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_000334135.1	Contig	437.106	CALO01	945079	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_001889245.1	Contig	2.42226	LOLC01	9834	0
Triticum aestivum	Eukaryota; Plants; Land Plants	GCA_001889205.1	Contig	0.942499	LOLD01	3746	0
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_900323645.1	Scaffold	392.684	OOGV01	252227	0
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_900323695.1	Scaffold	552.916	OOHG01	360242	0
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_900323615.1	Scaffold	477.083	OOGX01	350258	0
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_900323635.1	Scaffold	434.779	OOGY01	363434	0
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_900323595.1	Scaffold	431.061	OOHA01	427274	0
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_900323585.1	Scaffold	288.316	OOGW01	327364	0
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_900323625.1	Scaffold	518.951	OOHB01	592030	0
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_900323565.1	Scaffold	353.054	OOGZ01	455525	0
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_900323665.1	Scaffold	441.177	OOHC01	572876	0
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_900323605.1	Scaffold	497.42	OOHD01	649826	0
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_900323575.1	Scaffold	585.067	OOHE01	730770	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_900323655.1	Scaffold	624.303	OOHF01	905302	0
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_900323675.1	Scaffold	388.595	OOHH01	638342	0
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_900323685.1	Scaffold	700.417	OOHP01	1181661	0
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_002162155.2	Chromosome	10677.9	LSYQ02	148396	0
Triticum dicoccoides	Eukaryota; Plants; Land Plants	GCA_900184675.1	Chromosome	10495	FXXJ01	149145	0
Triticum urartu	Eukaryota; Plants; Land Plants	GCA_000347455.1	Scaffold	3747.05	AOTI01	499221	24169
Triticum urartu	Eukaryota; Plants; Land Plants	GCA_003073215.1	Chromosome	4851.9	MKGO01	10284	0
Turritis glabra	Eukaryota; Plants; Land Plants	GCA_900406565.1	Contig	171.13	OVBN01	250	0
Turritis glabra	Eukaryota; Plants; Land Plants	GCA_900406545.1	Scaffold	157.63	OVBM01	6441	0
Urochloa ruziziensis	Eukaryota; Plants; Land Plants	GCA_003016355.1	Scaffold	732.531	PVZT01	102577	0
Utricularia gibba	Eukaryota; Plants; Land Plants	GCA_002189035.1	Chromosome	100.689	NEEC01	518	0
Vaccinium macrocarpon	Eukaryota; Plants; Land Plants	GCA_000775335.2	Scaffold	414.622	JOTO01	200203	0
Vachellia collinsii	Eukaryota; Plants; Land Plants	GCA_006871305.1	Scaffold	461.065	QFDD01	122260	0
Vanilla planifolia	Eukaryota; Plants; Land Plants	GCA_004338375.1	Scaffold	2203.64	SDXO01	794534	0

Vicia faba	Eukaryota; Plants; Land Plants	GCA_001375635.1	Contig	80.3627	CSVX01	74659	0
Vigna angularis	Eukaryota; Plants; Land Plants	GCA_001190045.1	Chromosome	467.301	JZJH01	37727	37769
Vigna angularis var. angularis	Eukaryota; Plants; Land Plants	GCA_000465365.1	Scaffold	291.824	AUGG01	14501	0
Vigna angularis var. angularis	Eukaryota; Plants; Land Plants	GCA_001723775.1	Chromosome	444.439	JRFV01	3387	0
Vigna radiata	Eukaryota; Plants; Land Plants	GCA_000180895.1	Contig	10.1012	BABL01	46645	0
Vigna radiata var. radiata	Eukaryota; Plants; Land Plants	GCA_001584445.1	Scaffold	454.907	LJIH01	2418	0
Vigna radiata var. radiata	Eukaryota; Plants; Land Plants	GCA_000741045.2	Chromosome	463.638	JJMO01	2499	42284
Vigna unguiculata	Eukaryota; Plants; Land Plants	GCA_004118075.1	Chromosome	519.067	NBOW01	682	41173
Vigna unguiculata subsp. unguiculata	Eukaryota; Plants; Land Plants	GCA_001687525.1	Scaffold	695.046	MATU01	224035	0
Viola pubescens var. scabriuscula	Eukaryota; Plants; Land Plants	GCA_002752925.1	Scaffold	318.366	NBIL01	157716	0
Vitis aestivalis	Eukaryota; Plants; Land Plants	GCA_001562795.1	Contig	432.755	LOML01	756125	0
Vitis cinerea x Vitis riparia	Eukaryota; Plants; Land Plants	GCA_001282645.1	Scaffold	539.624	CCJE01	210444	0
Vitis riparia	Eukaryota; Plants; Land Plants	GCA_004353265.1	Chromosome	500.106	SJAQ01	174	0
Vitis vinifera	Eukaryota; Plants; Land Plants	GCA_002923105.1	Scaffold	427.211	BDSR01	21	0

(continued)

Table 32.3 (continued)

Organism name	Organism groups	Assembly	Level	Size (Mb)	WGS	Scaffolds	CDS
Vitis vinifera	Eukaryota; Plants; Land Plants	GCA_002922885.1	Scaffold	427.171	BDSO01	21	0
Vitis vinifera	Eukaryota; Plants; Land Plants	GCA_002923015.1	Scaffold	427.04	BDSQ01	21	0
Vitis vinifera	Eukaryota; Plants; Land Plants	GCA_002923165.1	Scaffold	426.616	BDSS01	21	0
Vitis vinifera	Eukaryota; Plants; Land Plants	GCA_004011995.1	Contig	868.043	QGNW01	2737	112320
Vitis vinifera	Eukaryota; Plants; Land Plants	GCA_000003745.2	Chromosome	486.197	CAAP03	2061	41208
Vitis x labruscana x Vitis vinifera	Eukaryota; Plants; Land Plants	GCA_008326845.1	Scaffold	490.143	BKBX01	8696	0
Xanthoceras sorbifolium	Eukaryota; Plants; Land Plants	GCA_003430845.1	Chromosome	504.383	QUWJ01	2297	0
Xerophyta viscosa	Eukaryota; Plants; Land Plants	GCA_002076135.1	Scaffold	295.462	MJHO01	896	0
Zea mays	Eukaryota; Plants; Land Plants	GCA_000223545.1	Scaffold	177.051	AECO01	196697	0
Zea mays	Eukaryota; Plants; Land Plants	GCA_000275765.1	Contig	1.33507	AHID01	1844	0
Zea mays	Eukaryota; Plants; Land Plants	GCA_003185045.1	Chromosome	2182.61	NCVQ01	2203	46530
Zea mays	Eukaryota; Plants; Land Plants	GCA_003704525.1	Chromosome	2198.5	RAQR01	797	0
Zea mays	Eukaryota; Plants; Land Plants	GCA_003709335.1	Chromosome	2288.19	RAQT01	972	0
Zea mays	Eukaryota; Plants; Land Plants	GCA_000005005.6	Chromosome	2135.08	LPUQ01	598	58411

Zea mays subsp. mays	Eukaryota; Plants; Land Plants	GCA_002813505.1	Scaffold	2041.55	LMUZ01	48268	0
Zea mays subsp. mays	Eukaryota; Plants; Land Plants	GCA_001990705.1	Chromosome	2392.8	MTTB01	62610	0
Zea mays subsp. mays	Eukaryota; Plants; Land Plants	GCA_001984235.2	Chromosome	2455.26	MTTA01	60567	0
Zea mays subsp. mays	Eukaryota; Plants; Land Plants	GCA_001644905.2	Chromosome	2133.88	LWRW02	191	0
Zea mays subsp. mays	Eukaryota; Plants; Land Plants	GCA_002682915.2	Chromosome	2197.97	NWUM01	3538	0
Zea mays subsp. mays	Eukaryota; Plants; Land Plants	GCA_002237485.1	Chromosome	2155.82	NKIA01	43301	0
Zea mays subsp. mexicana	Eukaryota; Plants; Land Plants	GCA_002813485.1	Scaffold	1204.28	LMVA01	107418	0
Zizania latifolia	Eukaryota; Plants; Land Plants	GCA_000418225.1	Scaffold	603.989	ASSH01	4522	0
Ziziphus jujuba	Eukaryota; Plants; Land Plants	GCA_001835785.1	Scaffold	351.097	LPXJ01	36119	0
Ziziphus jujuba	Eukaryota; Plants; Land Plants	GCA_000826755.1	Chromosome	437.754	JREP01	5897	43574
Zostera marina	Eukaryota; Plants; Land Plants	GCA_001185155.1	Scaffold	203.914	LFYR01	2228	20648
Zoysia japonica	Eukaryota; Plants; Land Plants	GCA_001602275.1	Scaffold	334.384	BCLF01	11786	0
Zoysia matrella	Eukaryota; Plants; Land Plants	GCA_001602295.1	Scaffold	563.439	BCLG01	13609	0
Zoysia pacifica	Eukaryota; Plants; Land Plants	GCA_001602315.1	Scaffold	397.01	BCLH01	11428	0

32.6 Conclusion

The applications of bioinformatics to plant pathology have been pivotal role in understanding of host and pathogen evolution and molecular interactions between host and pathogen. Availability of next-generation sequencing data of candidate model organisms of all kingdom through high-throughput technology is convenient to deal with biological systems and understand the biological sequence–structure–function correlation using in-silico biology tools, technology and databases. Genome annotation, assembly, bioproject, biosample submission, sequence data submission, retrieval of data, data analysis, variation analysis, conserved domain analysis, gene identification, regulatory elements analysis, gene expression analysis, structure prediction, structure visualization, structure analysis, structure classification, molecular modeling, epitope identification and mapping using 3D, drug designing, active site analysis and molecular docking, etc. play an important role to achieve biological function and understand the sequence–structure–function relationship. These all in-silico biology techniques will be further helpful in genomics-assisted crop improvement and development of designer crops with high yield and super quality.

References

Alioto T (2012) Gene prediction. Methods Mol Biol 855:175–201

Aljanabi S (2001) Genomics and plant breeding. Biotechnol Annu Rev 7:195–238

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Batley J, Edwards D (2016) The application of genomics and bioinformatics to accelerate crop improvement in a changing climate. Curr Opin Plant Biol 30:78–81

Bawono P, Dijkstra M, Pirovano W, Feenstra A, Abeln S, Heringa J (2017) Multiple sequence alignment. Methods Mol Biol 1525:167–189

Berman HM (2008) The protein data bank: a historical perspective. Acta Crystallogr A 64 (Pt 1):88–95

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28(1):235–242

Bolger ME, Arsova B, Usadel B (2018) Plant genome and transcriptome annotations: from misconceptions to simple solutions. Brief Bioinform 19:437–449

Bolser D, Staines DM, Pritchard E, Kersey P (2016) Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data. Methods Mol Biol 1374:115–140

Bolser DM, Staines DM, Perry E, Kersey PJ (2017) Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomic data. Methods Mol Biol 1533:1–31

Cavasotto CN, Phatak SS (2009) Homology modeling in drug discovery: current trends and applications. Drug Discov Today 14:676–683

Chen YP, Chen F (2008) Using bioinformatics techniques for gene identification in drug discovery and development. Curr Drug Metab 9:567–573

Chen J, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, Hurwitz DI, Jackson JD, Jacobs AR, Lanczycki CJ, Liebert CA, Liu C, Madej T, Marchler-Bauer A, Marchler GH, Mazumder R, Nikolskaya AN, Rao BS, Panchenko AR, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Vasudevan S, Wang Y, Yamashita RA, Yin JJ, Bryant SH (2003) MMDB: Entrez's 3D-structure database. Nucleic Acids Res 31:474–477

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31:3497–3500

Clough E, Barrett T (2016) The gene expression omnibus database. Methods Mol Biol 1418:93–110

De Las Rivas J, Fontanillo C (2010) Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. PLoS Comput Biol 6(6):e1000807

de Ruyck J, Brysbaert G, Blossey R, Lensink MF (2016) Molecular docking as a popular tool in drug design, an in silico travel. Adv Appl Bioinforma Chem 9:1–11

Demuth JP, Hahn MW (2009) The life and death of gene families. BioEssays 31:29–39

Dong Q, Schlueter SD, Brendel V (2004) PlantGDB, plant genome database and analysis tools. Nucleic Acids Res 32(Database issue):D354–D359

Duvick J, Fu A, Muppirala U, Sabharwal M, Wilkerson MD, Lawrence CJ, Lushbough C, Brendel V (2008) PlantGDB: a resource for comparative plant genomics. Nucleic Acids Res 36(Database issue):D959–D965

Eisenberg D, Lüthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 277:396–404

Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD (2015) Molecular docking and structure-based drug design strategies. Molecules 20:13384–13421

França TC (2015) Homology modeling: an important tool for the drug discovery. J Biomol Struct Dyn 33(8):1780–1793

Galperin MY, Fernández-Suárez XM (2012) The 2012 nucleic acids research database issue and the online molecular biology database collection. Nucleic Acids Res 40(Database issue):D1–D8

Gebhardt C, Schmidt R, Schneider K (2005) Plant genome analysis: the state of the art. Int Rev Cytol 247:223–284

Gonçalves JP, Madeira SC, Oliveira AL (2009) BiGGEsTS: integrated environment for biclustering analysis of time series gene expression data. BMC Res Notes 2:124

Goodsell DS, Morris GM, Olson AJ (1996) Automated docking of flexible ligands: applications of AutoDock. J Mol Recognit 9:1–5

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40(Database issue):D1178–D1186

Gschwend DA, Good AC, Kuntz ID (1996) Molecular docking towards drug discovery. J Mol Recognit 9:175–186

Guedes IA, de Magalhães CS, Dardenne LE (2014) Receptor-ligand molecular docking. Biophys Rev 6:75–87

Hardison RC (2003) Comparative genomics. PLoS Biol 1:E58

Higgins DG, Thompson JD, Gibson TJ (1996) Using CLUSTAL for multiple sequence alignments. Methods Enzymol 266:383–402

Jones G, Willett P (1995) Docking small-molecule ligands into active sites. Curr Opin Biotechnol 6:652–656

King GJ (2004) Bioinformatics: harvesting information for plant and crop science. Semin Cell Dev Biol 15:721–731

Kodama Y, Mashima J, Kaminuma E, Gojobori T, Ogasawara O, Takagi T, Okubo K, Nakamura Y (2012) The DNA Data Bank of Japan launches a new resource, the DDBJ Omics Archive of functional genomics experiments. Nucleic Acids Res 40(Database issue):D38–D42

Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S (2017) The ClusPro web server for protein-protein docking. Nat Protoc 12:255–278

Krieger E, Vriend G (2014) YASARA view – molecular graphics for all devices – from smartphones to workstations. Bioinformatics 30:2981–2982

Krieger E, Nabuurs SB, Vriend G (2003) Homology modeling. Methods Biochem Anal 44:509–523

Kroemer RT (2007) Structure-based drug design: docking and scoring. Curr Protein Pept Sci 8:312–328

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35(6):1547–1549

Kuntal BK, Aparoy P, Reddanna P (2010) Easy modeller: a graphical interface to MODELLER. BMC Res Notes 3:226

Laskowski RA, Hutchinson EG, Michie AD, Wallace AC, Jones ML, Thornton JM (1997) PDBsum: a web-based database of summaries and analyses of all PDB structures. Trends Biochem Sci 22(12):488–490. PubMed PMID: 9433130

Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM (2017) PDBsum: structural summaries of PDB entries. Protein Sci 27:129–134

Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins 50:437–450

Lybrand TP (1995) Ligand-protein docking and rational drug design. Curr Opin Struct Biol 5:224–228

Lyons E, Freeling M (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J 53:661–673

Macindoe G, Mavridis L, Venkatraman V, Devignes MD, Ritchie DW (2010) HexServer: an FFT-based protein docking server powered by graphics processors. Nucleic Acids Res 38 (Web Server issue):W445–W449

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47(W1):W636–W641

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45:D200–D203

Martinez M (2013) From plant genomes to protein families: computational tools. Comput Struct Biotechnol J 8:e201307001

McClure MA, Vasi TK, Fitch WM (1994) Comparative analysis of multiple protein-sequence alignment methods. Mol Biol Evol 11:571–592

Meng XY, Zhang HX, Mezei M, Cui M (2011) Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des 7:146–157

Metsalu T, Vilo J (2015) ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. Nucleic Acids Res 43:W566–W570

Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F, Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G, Orengo C, Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND, Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-Vegas A, Sigrist CJA, Sillitoe I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE, Yong SY, Finn RD (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res 47:D351–D360

Mochida K, Shinozaki K (2010) Genomics and bioinformatics resources for crop improvement. Plant Cell Physiol 51:497–523

Moody G (2004) Digital code of life: how bioinformatics is revolutionizing science, medicine, and business. Wiley, Hoboken. ISBN 978-0-471-32788-2

Morris GM, Lim-Wilby M (2008) Molecular docking. Methods Mol Biol 443:365–382

Mount DW (2007) Using the basic local alignment search tool (BLAST). CSH Protoc 2007:pdb. top17

NCBI Resource Coordinators (2016) Database resources of the national center for biotechnology information. Nucleic Acids Res 44:D7–D19

Ong Q, Nguyen P, Thao NP, Le L (2016) Bioinformatics approach in plant genomic research. Curr Genomics 17:368–378

Pagadala NS, Syed K, Tuszynski J (2017) Software for molecular docking: a review. Biophys Rev 9:91–102

Pirovano W, Heringa J (2008) Multiple sequence alignment. Methods Mol Biol 452:143–161. https://doi.org/10.1007/978-1-60327-159-2_7. Review. PubMed PMID: 18566763

Rambaut A (2012) FigTree v1. 4.0. University of Oxford, Oxford. http://tree.bio.ed.ac.uk/software/figtree/

Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, Connor R, Fiorini N, Funk K, Hefferon T, Holmes JB, Kim S, Kimchi A, Kitts PA, Lathrop S, Lu Z, Madden TL, Marchler-Bauer A, Phan L, Schneider VA, Schoch CL, Pruitt KD, Ostell J (2019) Database resources of the national center for biotechnology information. Nucleic Acids Res 47:D23–D28

Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res 33(Web Server issue):W363–W367

Seo J, Gordish-Dressman H, Hoffman EP (2006) An interactive power analysis tool for microarray hypothesis testing and generation. Bioinformatics 22(7):808–814

Sievers F, Higgins DG (2014) Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol Biol 1079:105–116

Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P (2002) PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform 3(3):265–274

Sousa SF, Fernandes PA, Ramos MJ (2006) Protein-ligand docking: current status and future challenges. Proteins 65:15–26

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 47:D607–D613

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics. Chapter 2:Unit 2.3

Tippmann HF (2004) Analysis for free: comparing programs for sequence analysis. Brief Bioinform 5(1):82–87

Trosset JY, Cavé C (2019) In silico drug-target profiling. Methods Mol Biol 1953:89–103

UniProt Consortium T (2018) UniProt: the universal protein knowledgebase. Nucleic Acids Res 46:2699

Wang Z, Chen Y, Li Y (2004) A brief review of computational gene prediction methods. Genomics Proteomics Bioinformatics 2:216–221

Wheeler DL, Smith-White B, Chetvernin V, Resenchuk S, Dombrowski SM, Pechous SW, Tatusova T, Ostell J (2005) Plant genome resources at the national center for biotechnology information. Plant Physiol 138:1280–1288

Xiang Z (2006) Advances in homology protein structure modeling. Curr Protein Pept Sci 7:217–227. Review. PubMed PMID: 16787261; PubMed Central PMCID: PMC1839925

Xu D, Xu Y, Uberbacher EC (2000) Computational tools for protein modeling. Curr Protein Pept Sci 1:1–21

Zhao H, Caflisch A (2015) Molecular dynamics in drug design. Eur J Med Chem 91:4–14

URLs

http://bioinformatics.cau.edu.cn/PMRD/
http://circos.ca/
http://hex.loria.fr/
http://mordred.bioc.cam.ac.uk/~rapper/rampage.php
http://tree.bio.ed.ac.uk/software/figtree/
http://www.cs.umd.edu/hcil/hce/
http://www.mbio.ncsu.edu/BioEdit/bioedit.html
http://www.plantgdb.org/
https://biit.cs.ut.ee/clustvis/
https://bioinfo3d.cs.tau.ac.il/PatchDock/
https://blast.ncbi.nlm.nih.gov/Blast.cgi
https://en.wikipedia.org/wiki/Category:Bacterial_plant_pathogens_and_diseases
https://en.wikipedia.org/wiki/Category:Fungal_plant_pathogens_and_diseases
https://en.wikipedia.org/wiki/Category:Lists_of_plant_diseases
https://en.wikipedia.org/wiki/Category:Viral_plant_pathogens_and_diseases
https://phytozome.jgi.doe.gov/pz/portal.html
https://plants.ensembl.org/index.html
https://prosite.expasy.org/
https://servicesn.mbi.ucla.edu/Verify3D/
https://www.ddbj.nig.ac.jp/index-e.html
https://www.ebi.ac.uk/interpro/search/sequence/
https://www.ebi.ac.uk/Tools/msa/clustalo/
https://www.embl.org/
https://www.megasoftware.net/index.html
https://www.ncbi.nlm.nih.gov/
https://www.ncbi.nlm.nih.gov/geo/
https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml
https://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml
https://www.plantcyc.org/
https://www.rcsb.org/
https://www.uniprot.org/