

# Genome Assembly

- [Reference Genome 2024](#)
- [Reference Genome 2000-2023](#)
  - [Known Gaps in the 2000 Chromosome Assembly](#)
  - [GFF file of all known gaps in the Arabidopsis genome assembly April 2008](#)
  - [Clones Missing or Incomplete in GenBank September 2003](#)
  - [Table of Gaps and Incomplete Clones September 2003](#)
  - [AGI Groups](#)

## Reference Genome 2024

Progress on the newest reannotation can be found [here](#).

## Reference Genome 2000-2023


The Arabidopsis thaliana genome was sequenced in 2000 by the Arabidopsis Genome Initiative (AGI) ([Nature 14 Dec. 2000](#)). The genome has five chromosomes and a total size of approximately 135-megabases. The current TIGR golden path length is 119,146,348 bp. The table below shows the approximate total length and the length of the golden path for each chromosome.

	Golden path length	Approximate chromosome length
Chromosome 1	30,427,671 bp	34,964,571 bp
Chromosome 2	19,698,289 bp	22,037,565 bp
Chromosome 3	23,459,830 bp	25,499,034 bp
Chromosome 4	18,585,056 bp	20,862,711 bp
Chromosome 5	26,975,502 bp	31,270,811 bp
Total	119,146,348 bp	134,634,692 bp

[Chromosome sequence data](#) and [AGI tiling paths](#) are available from the [TAIR Downloads](#).

[TAIR8\\_Assembly\\_updates.xls](#) and [TAIR9\\_Assembly\\_updates.xls](#) contains a list of all assembly updates made for the TAIR8 and TAIR9 genome releases.

## Known Gaps in the 2000 Chromosome Assembly



Centromeres and other gaps between clones in **red**.

Clones containing gaps in **purple**.

\* Indicates sequence is not yet deposited in GenBank.

### Chromosome 1:

T18N24-**F8L2**-F2C1-**F12G6**-T23P23-**T28N5**-F11K13

T24F19-**CEN1**-F13P3

F9A12-**F25O15**-**F9D18**-T5F23

F27F5-**T2P3**-F2G19

F12A4-**F1504**-F14D7

T32E22-**F103**-T32E20

F16N3-**T2E6**-T6B12

F10A5-**T4012**-T23E18

### Chromosome 2:

NOR2-F23H14-F10A8

T12J2-CEN2-T6C20-T14C8

T4E5-F10C8-T18E17

#### Chromosome 3:

TEL3N-T4P13

K3G3-MJL12-MTE24

MUO10-T13B17-MWE13

F8N14-T803-F1M23

T15D2-CEN3-T25F15-F23H6-T28G19-5SrDNA-F1C23-T18B3-T26P13-T14A11-T4P3-F21A14-5SrDNA-F4M19 F7M19-T6L19- -F7K15

#### Chromosome 4:

NOR4-T15P10 F21I2-5SrDNA-F14G16

T2N12-CEN4-F13J5

T13J8-F26K10-F20O9-T5F17-F16A16

F19B15-F17A13-T16L4

F6I18-F6E21-F8F16

F4D11-T16I18-F26P21

#### Chromosome 5:

F21E1-T19N18\*-T32M21

F23C8-T26N4-5SrDNA-F23B23

F28N5-CEN5-T8H11

T32B3-5SrDNA-T25B21-T3J11

[GFF file of all known gaps in the Arabidopsis genome assembly](#) April 2008

[Clones Missing or Incomplete in GenBank](#) September 2003

Clones in GenBank HTG section (sequencing in progress) or missing from GenBank. Includes chromosome, status, accession number, group and comments.

### Table of Gaps and Incomplete Clones September 2003

Includes comments from TAIR, TIGR and AGI groups on status and priority for sequencing.

Clone or Gap	TAIR Comment	TIGR Comment	Other Comment
<b>Chromosome 1</b>			
Clone F8L2		CEN1: 3 asm; 3 grps. Will most likely close but mainly Athila etc. Priority 3.	
Clone F12G6	GenBank PLN, centromeric, 3 unordered (SSP)		Submitted by Ecker
Clone T28N5	GenBank PLN, centromeric, 11 unordered (SSP)		Submitted by Ecker
Clone F25O15		CEN1: 2 asm; 1 grp. This BAC should close soon. Priority 3.	

Clone F9D18	GenBank PLN, centromeric, contains gap of approx. 34 kb (SSP)		Theologis group has updated GenBank record
Clone T2P3		Euchr 1: ~20 asm; 10 grps. Contains >99% identical repeats. We have made many different libraries for this BAC in attempts to get more uniform coverage. Two ~30 kb and ~ 35 kb assemblies that overlap the neighboring BACs have been annotated to assess their gene content. Transposon-based sequencing of linking clones continues. We expect to be able to reduce the number of assemblies and annotate then even if full closure is not possible. Priority 1.	
<b>Chromosome 2</b>			
Clone T6C20		CEN2: This BAC lies within the 180 bp repeats; thus full closure was not attempted in 1998-1999 when chromosome 2 was completed. We have revisited this BAC, including making a large insert library, but find that this does not help. This BAC will not be closed. Priority 4.	
Clone F10C8	GenBank PLN, contains gap of approx. 20 kb consisting of 747 bp pure tandem repeats		
<b>Chromosome 3</b>			
Clone T25F15		CEN3: This BAC lies in one of two small island contigs within CEN3 and contains 180 bp repeats. It is in 2 asm, 1 grp with one apparent sequencing gap. Nested deletions of a spanning clone have been made that should in theory close this BAC and are in the sequencing queue. Priority 3.	
Clone T28G19		CEN3: 22 asms; 10 grps even when edited to 6+ and assembled at 100%. This BAC lies in one of two small island contigs within CEN3. Most likely the status of this BAC cannot be improved by further effort. Priority 4.	
Clone T18B3		CEN3: 8 asms; 4 grps. We can make progress with this BAC and will do so. Priority 3.	
Clone F21A14		CEN3: 26 asms; 3 grps after editing and assembly at 100%. This BAC cannot be closed. Priority 4.	
<b>Chromosome 4</b>			
Clone F26K10	HTG (Sanger)		
Clone T5F17	HTG (Sanger)		
Clone F17A13	HTG (Sanger)		
Clone F6E21	HTG (Sanger)		
Clone T16I18	HTG (Sanger)		
<b>Chromosome 5</b>			
Clone T19N18	No GenBank record (Sanger?) TIGR has annotated sequence for this clone.		
Clone T26N4	HTG, 15 unordered pieces		CSHL finishing (includes T22F2). Contains 5S rRNA repeats.
Clone T25B21	HTG, 3 unordered pieces		CSHL finishing, contains 5S rRNA repeats.
<b>Unanchored Clones</b>			
F26J21		65 asms; 32 gps (but most sequences fall into 5 asms, 2 grps). Appears to contain many genes. We plan to close and resolve as much as possible. Awaiting more sequence. Priority 2.	
T13I7		21 asms; 9 grps. We plan to close and resolve as much as possible. Awaiting large insert library and more sequence. Priority 2.	

T4N13		52 asms; 37 grps. We plan to close and resolve as much as possible. Awaiting large insert library and more sequence. Priority 2.	
T17J15		26 asms; 14 grps. We plan to close and resolve as much as possible. Awaiting large insert library and more sequence. Priority 2.	

## AGI Groups

### Cold Spring Harbor Sequencing Consortium (CSHSC)

<b>Members:</b>	<a href="#">CSHL</a> , <a href="#">ABI</a> , <a href="#">WashU</a>
<b>Contacts:</b>	Dick McCombie, Rob Martienssen (CSHL); Rick Wilson (WashU)
<b>Regions sequenced:</b>	13.1 Mb including the top of chromosome 4 and 3 Mb around the centromere of chromosome 5.

### European Scientists Sequencing Arabidopsis (ESSA)

<b>Members:</b>	<a href="#">John Innes Centre</a> , MIPS, network of 18 labs
<b>Contacts:</b>	Mike Bevan (JIC); Klaus Mayer (MIPS)
<b>Regions sequenced:</b>	Chromosomes 4 (14.5 Mb) and 5 (6Mb)

### Genoscope-EU Consortium

<b>Members:</b>	EMBL, <a href="#">Genoscope</a> , Lion, U. van Amsterdam, Valle
<b>Contacts:</b>	Marcel Salanoubat, Francis Quetier
<b>Regions sequenced:</b>	Chromosome 3 bottom arm (9.2 Mb)

### Kazusa DNA Research Institute

<b>Members:</b>	<a href="#">Kazusa</a>
<b>Contacts:</b>	Satoshi Tabata, Kiyotaka Okada
<b>Regions sequenced:</b>	Chromosomes 3 (9.8 Mb) and 5 (17.8 Mb)

### SPP Consortium

<b>Members:</b>	PGEC, Stanford, UPenn (ATGC)
<b>Contacts:</b>	Sakis Theologis (PGEC); Ron Davis (Stanford); Joe Ecker (ATGC)
<b>Regions sequenced:</b>	Chromosome 1 (20.2 Mb)

### The J. Craig Venter Center (JCVI) formerly TIGR

<b>Members:</b>	<a href="#">JCVI</a>
<b>Contacts:</b>	Christopher Town
<b>Regions sequenced:</b>	Chromosome 2 (19.6 Mb), parts of 1 and 3