

The Celera WGS Assembler

<u>3-code</u>	<u>Message name</u>	<u>Meaning</u>	<u>Contents</u>
A "pull" from Content Systems			
frg	Input filename		
FRG	Fragment	Read or guide	Sequence, quality, type, source
DST	Distance	Len between pairs	Mean & std
LKG	Link	Mate pair info	2 frags, 1 dist, orientation, type
ADT	Audit	Batch header	name, time, version, comment
scn	Input filename		
SCN	Screen item	Known repeat	Sequence, category, hard/soft
RPT	Repeat item	Repeat category	Name, length
GateKeeper			
inp	Output filename		
IFG	Internal fragment	Fragment, internal ID	FRG with converted IDs
IDT	Internal distance	Distance, internal ID	DST with converted IDs
ILK	Internal link	Link, internal ID	LKG with converted IDs
ISC	Internal screen	Screen, internal ID	SCN with converted IDs
RPT	Repeat item	Repeat category	Name, length
Screen index file?			
Repeat Tagger & Contaminant Screener			
urc	Output filename		
SFG	Screened fragment	Masked fragment	IFG, clear range, alignment list
? ISM	Screen matches	Ungapped alignment	other id, 2 intervals, orientation
<i>IDT</i>	<i>Internal distance</i>	<i>Link, internal ID</i>	<i>DST with converted IDs</i>
<i>ILK</i>	<i>Internal link</i>	<i>Screen, internal ID</i>	<i>LKG with converted IDs</i>
<i>RPT</i>	<i>Repeat item</i>	<i>Repeat category</i>	<i>Name, length</i>
Overlapper / Fragment Graph Builder (FGB)			
All-against-all comparisons			
Writes to screened fragments store			
Reads & writes overlap store (for incremental restarts)			
ovl	Output filename		
OFG	Overlapped fragment	Fragment reference	SFG minus sequence & quality
OVL	Overlap	Pairwise alignment	2 frags, orientation, delta
<i>IDT</i>	<i>Internal distance</i>	<i>DST plus internal ID</i>	<i>All IDs converted to internal</i>
<i>ILK</i>	<i>Internal link</i>	<i>LKG plus internal ID</i>	<i>All IDs converted to internal</i>
<i>RPT</i>	<i>Repeat item</i>	<i>Repeat category</i>	<i>Name, length</i>
Unitigger / Chunk Graph Builder (CGB)			
Construct unitigs (maximal overlap graphs)			
Reads from screened fragments store			
cgb	Output filename		
IUM	Unitig	Vertices of graph	Consensus sequence & quality
IMP	Multi-position	Frag-to-unitig alignment	other id, 1 interval, delta
UOM	Unitig overlap	Edges of graph	2 unitigs, orientation, length
<i>IDT</i>	<i>Internal distance</i>	<i>DST plus internal ID</i>	<i>All IDs converted to internal</i>
<i>ILK</i>	<i>Internal link</i>	<i>LKG plus internal ID</i>	<i>All IDs converted to internal</i>
<i>RPT</i>	<i>Repeat item</i>	<i>Repeat category</i>	<i>Name, length</i>

Scaffolder / Chunk Graph Walker (CGW)

Construct scaffolds (maximal graphs)

Reads from screened fragments store

cgw Output filename

Build Extended Unitig Graph (EUG) using mate pairs as edges

IUM *Unitig* *Vertices of graph* *Consensus sequence & quality*

IUL Unitig link Edge of graph List of mate pairs

Build Extended Contig Graph (ECG) with unitigs and surrogates

PCM Pre-contig Contig cluster List of frag & unitig intervals

IEP Element position Unitig-to-contig alignment Interval of frag or unitig

ICL Contig link Edge of graph 2 contigs, orientation, type

IMD Mate distribution Distance analysis Distance ID, statistics

IAF Augmented fragment Fragment analysis Q/A on chimera, mate, clear rar

ISF Scaffold Best path thru contigs List of contig pairs

ICP Contig pair Internal to ISF 2 contigs, orientation, length

Consensus

Construct multi-seq alignments, fill gaps with repeat-unitigs

Reads from screened fragments store

con Output filename

ICM Contig Consensus Multi-seq alignment Consensus & quality, list of IMF

Terminator

Prepare external output

Reads from screened fragments store

asm Output filename

AFG Augmented fragment Genome snapshot

UTG

ULK

CCO

CLK

SCF

MDI

MPS

EPS

CTP