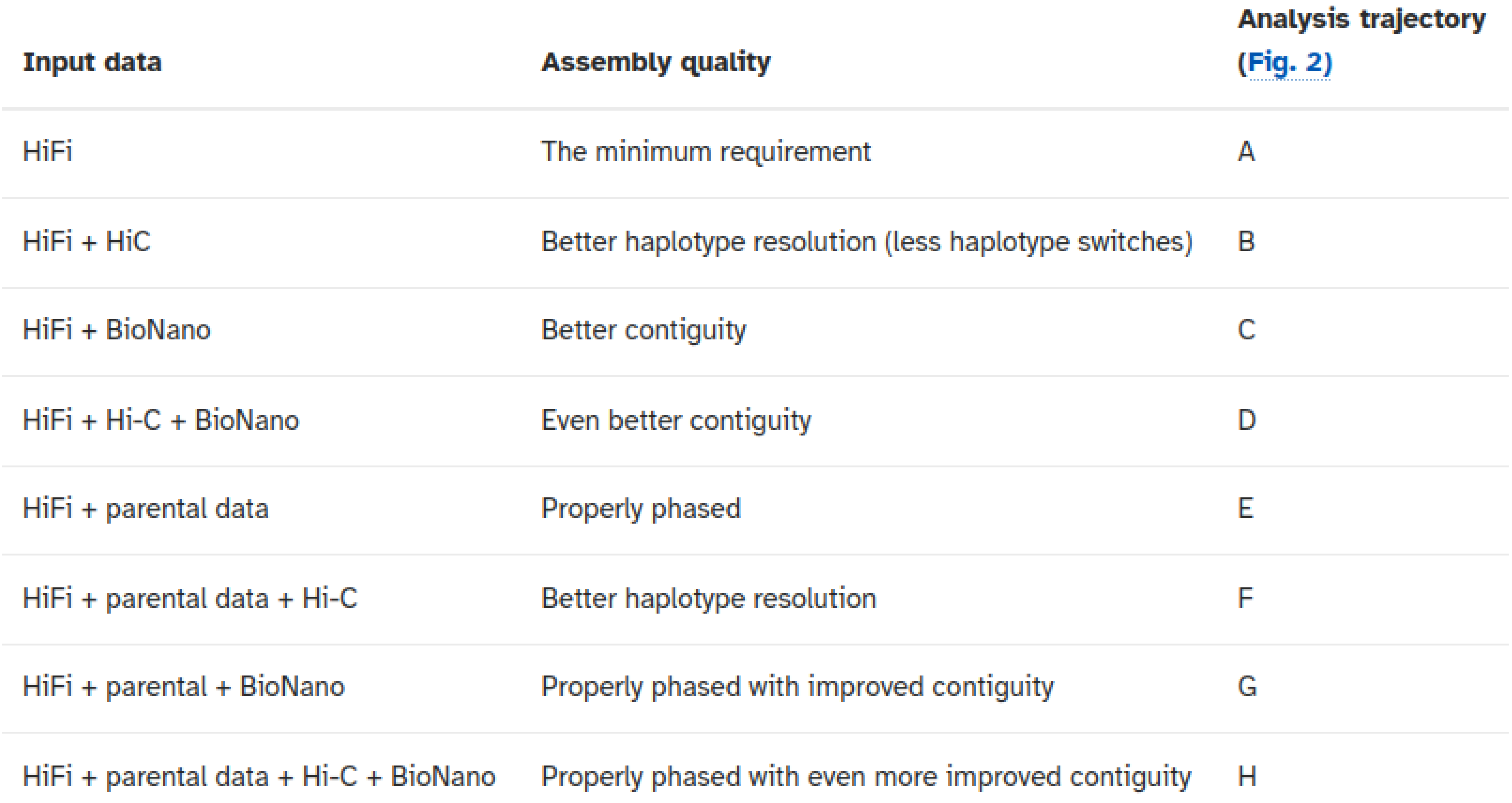
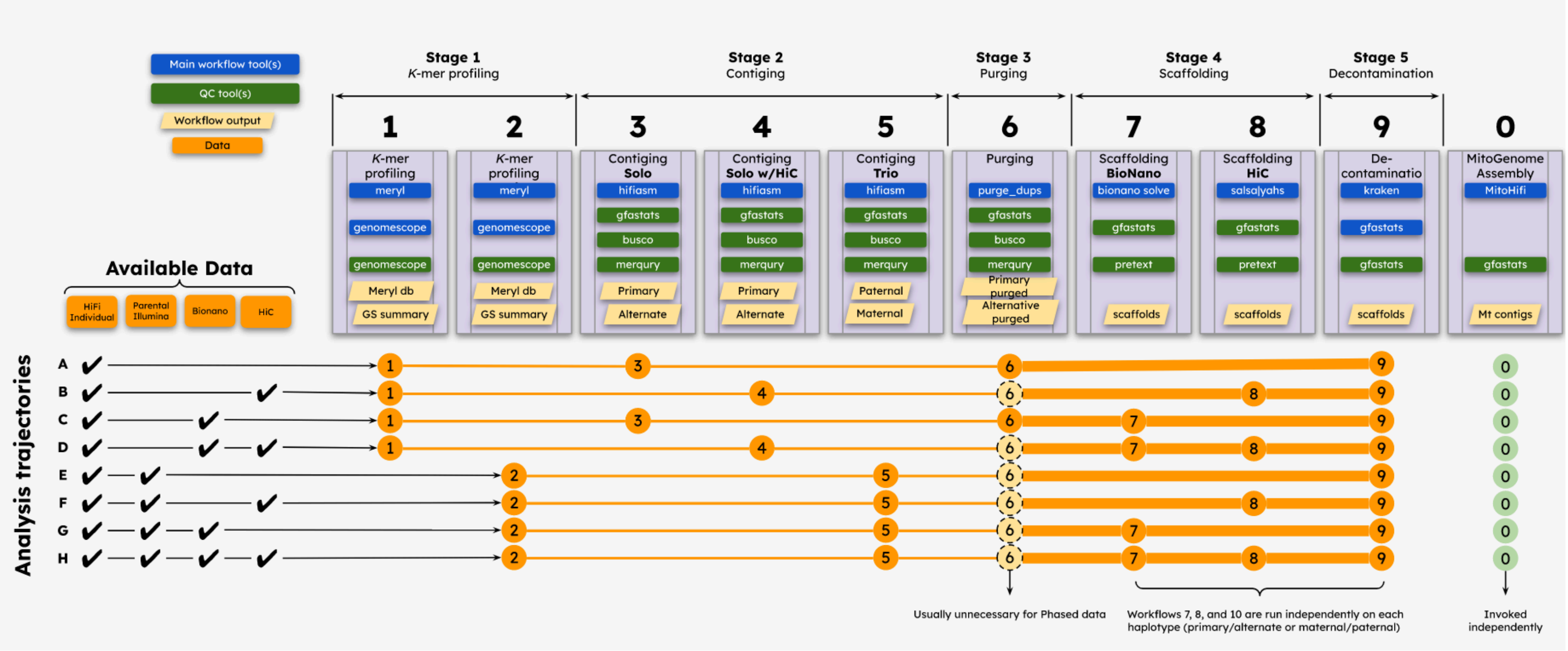
	Short reads only	Nanopore, CLR + short reads	HiFi (CCS), Nanopore (duplex)	
Adapted to	prokaryotic and very small eukaryotic genomes	any size but genomes with not too redundant repeats	any size	
read coverage	100x	50x + 50x	20x per haplotype	
read quality control	fastqc, jellyfish or kmc + genomescope2	fastqc, jellyfish or kmc + genomescope2	fastqc, jellyfish or kmc + genomescope2	
read cleaning software packages	fastp	Porechop (Nanopore) + fastp	HiFiAdapterFilt	
assembly	spades, megahit	NextDeNovo, flye	hifiasm	
polishing software packages	no polishing	medaka, racon for long reads, hapo-G for short reads	no polishing	
QC : assembly metrics	quast	quast	quast	
QC : nucleotide quality	meryl + merqury	meryl + merqury	meryl + merqury	
QC : kmer content	kat, merqury	kat, merqury	kat, merqury	
QC : gene content	BUSCO	BUSCO	BUSCO	
QC : comparison to reference genome	D-genies,Chromeister	D-genies,Chromeister	D-genies,Chromeister	
QC : contamination	blobtools	blobtools	blobtools	
Scaffolding	beware short contigs can	juicer + 3D-DNA, Yahs	juicer + 3D-DNA, Yahs	

	Short reads only	**Nanopore, CLR** **+ short reads**	**HiFi (CCS), Nanopore (duplex)**	
	be wrongly placed and oriented or not included in chromosomes juicer + 3D-DNA, Yahs			
Scaffolding manual curation	Juicebox, Pretextview	Juicebox, Pretextview	Juicebox, Pretextview	

VGP recipes for HiFi genome assembly
https://training.galaxyproject.org/training-material/topics/assembly/tutorials/vgp_genome_assembly/tutorial.html



Input data	**Assembly quality**	**Analysis trajectory (Fig. 2)**
HiFi	The minimum requirement	A
HiFi + HiC	Better haplotype resolution (less haplotype switches)	B
HiFi + BioNano	Better contiguity	C
HiFi + Hi-C + BioNano	Even better contiguity	D
HiFi + parental data	Properly phased	E
HiFi + parental data + Hi-C	Better haplotype resolution	F
HiFi + parental + BioNano	Properly phased with improved contiguity	G
HiFi + parental data + Hi-C + BioNano	Properly phased with even more improved contiguity	H


Main workflow tool(s)
QC tool(s)
Workflow output
Data
Stage 1
K-mer profiling
Stage 2
Contiging
Stage 3
Purging
Stage 4
Scaffolding
Stage 5
Decontamination
1
K-mer profiling
meryl
genomescope
genomescope
Meryl db
GS summary
2
K-mer profiling
meryl
genomescope
genomescope
Meryl db
GS summary
3
Contiging Solo
hifiasm
gfastats
busco
merqury
Primary
Alternate
4
Contiging Solo w/HiC
hifiasm
gfastats
busco
merqury
Primary
Alternate
5
Contiging Trio
hifiasm
gfastats
busco
merqury
Paternal
Maternal
6
Purging
purge_dups
gfastats
busco
merqury
Primary purged
Alternative purged
7
Scaffolding BioNano
bionano solve
gfastats
pretext
scaffolds
8
Scaffolding HiC
salsa|yahs
gfastats
pretext
scaffolds
9
De-contaminatio
kraken
gfastats
gfastats
scaffolds
0
MitoGenome Assembly
MitoHifi
gfastats
Mt contigs
Available Data
HiFi Individual
Parental Illumina
Bionano
HiC
Analysis trajectories
A
B
C
D
E
F
G
H
Usually unnecessary for Phased data
Workflows 7, 8, and 10 are run independently on each haplotype (primary/alternate or maternal/paternal)
Invoked independently