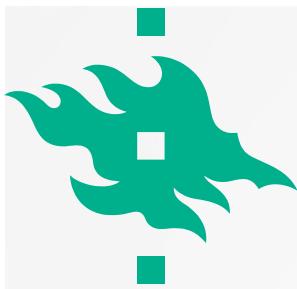




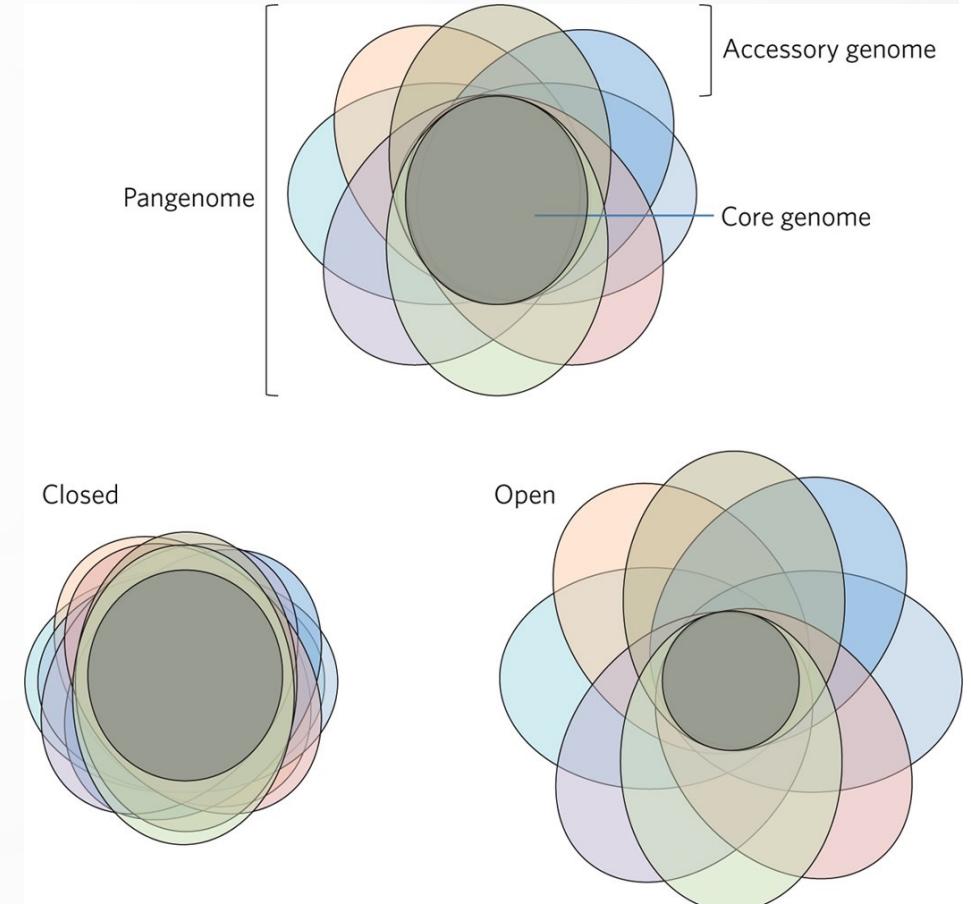
PANGENOMICS

(with anvi'o)

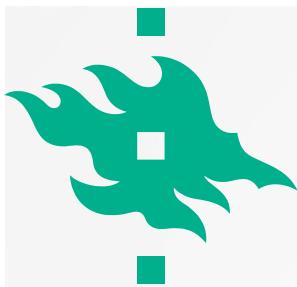


WHAT IS A PANGENOME?

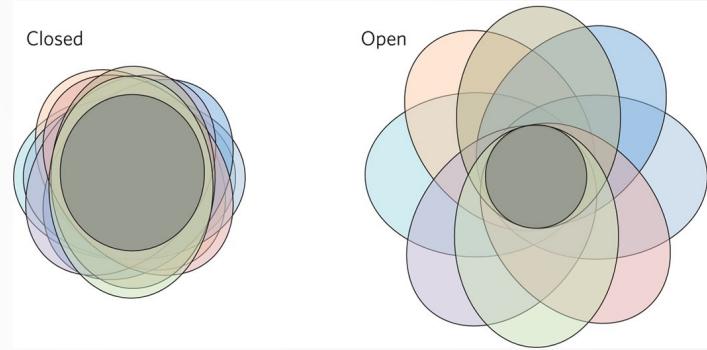
- Collection of genes found in a collection of genomes
- Core genome – genes shared by all genome
- Accessory genome – strain specific genes
- Closed or open pangenome



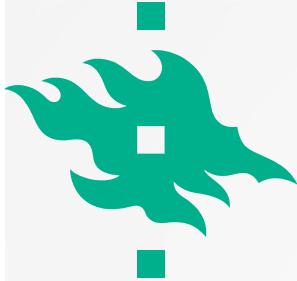
McInerney, J., McNally, A. & O'Connell, M. Why prokaryotes have pangenomes. *Nat Microbiol* 2, 17040 (2017). <https://doi.org/10.1038/nmicrobiol.2017.40>



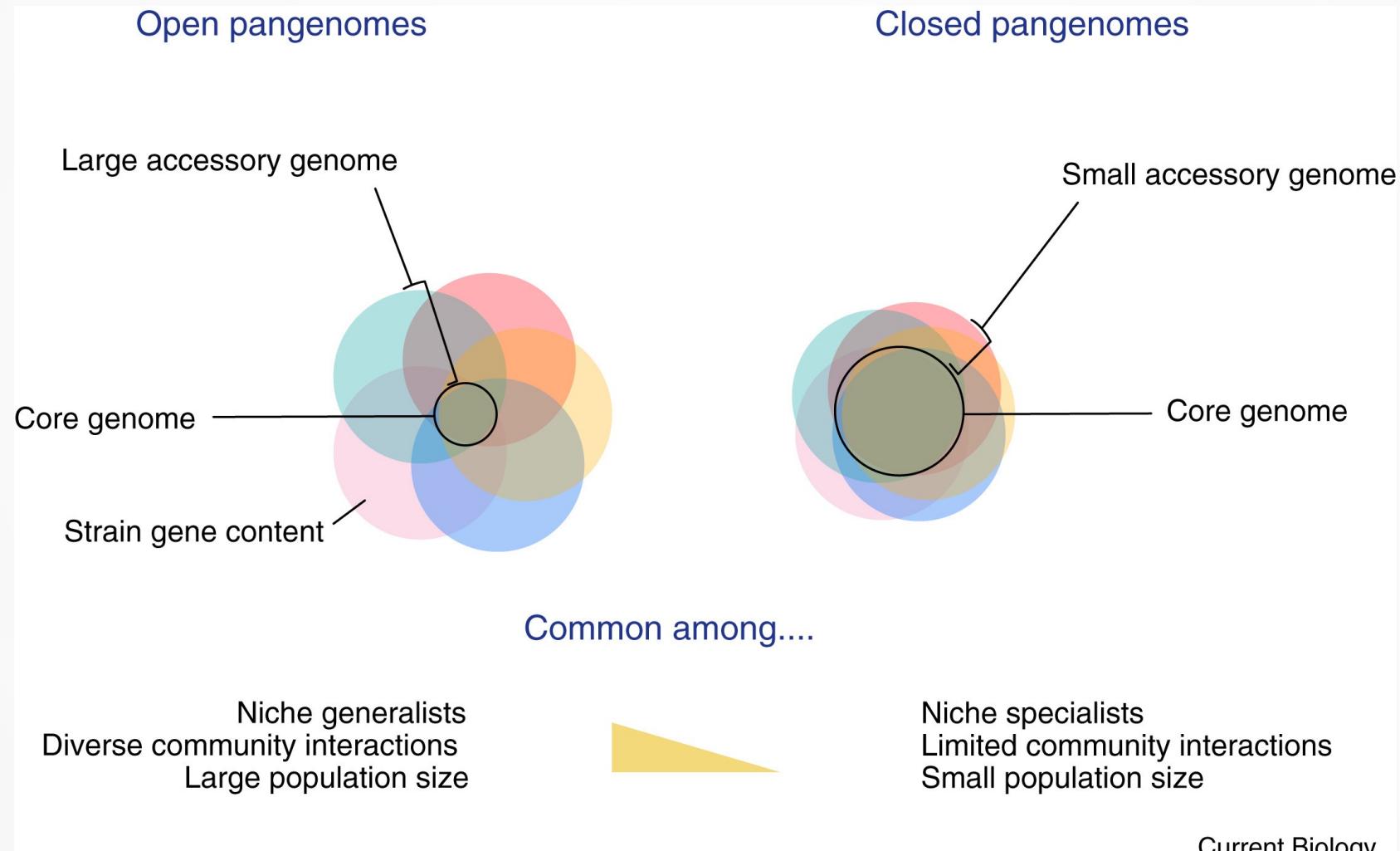
OPEN OR CLOSED PANGENOME?

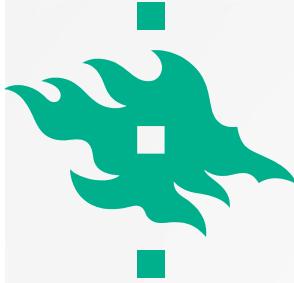


Species	Number of genomes	Core genome (% of total)	Accessory genome (% of total)	Pangenome size	No. of new genes per genome* (% of core)	Lifestyle
<i>Chlamydia trachomatis</i>	67	821 (84)	153 (15)	974	2 (0.2)	Obligate intracellular
<i>Bacillus anthracis</i>	42	4,001 (65)	2,065 (35)	6,066	49 (1.2)	Obligate pathogen
<i>Corynebacterium pseudotuberculosis</i>	40	1,205 (53)	1,041 (46)	2,246	26 (2.5)	Facultative intracellular pathogen
<i>Streptococcus pyogenes</i>	49	986 (37)	1,644 (62)	2,683	34 (3.4)	Human-restricted pathogen
<i>Staphylococcus aureus</i>	74	1,639 (36)	2879 (64)	4,52	39 (2.3)	Commensal with humans and other animals
<i>Corynebacterium diphtheriae</i>	13	1,632 (34)	3,156 (66)	4,786	243 (14.9)	Free-living
<i>Helicobacter pylori</i>	83	708 (28)	1,770 (71)	2,478	21 (3)	Intercellular, mucus-associated pathogen
<i>Acinetobacter baumannii</i>	40	1,997 (28)	5,042 (71)	7,039	126 (6.3)	Soil-dweller; opportunistic pathogen
						Associated with soil, rumen, guts of mammals and insects; occasionally pathogenic
<i>Bacillus cereus</i>	30	3,245 (27)	8,666 (72)	11,911	288 (8.9)	
<i>Klebsiella pneumoniae</i>	63	3,062 (26)	8,456 (74)	11,518	134 (4.4)	Free-living, azotroph, pathogen
<i>Prochlorococcus marinus</i>	40	1,175 (18)	5,260 (81)	6,435	132 (11.2)	Photosynthetic picoplankton
<i>Salmonella enterica</i>	206	1,717 (16)	8,550 (83)	10,267	42 (2.4)	Associated with rumen, water, soil; pathogen

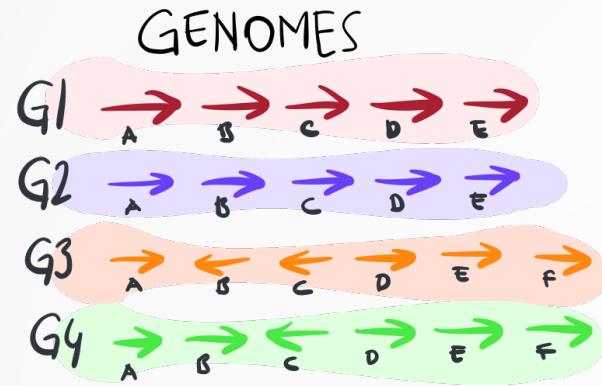


DRIVERS OF PANGENOME SIZE



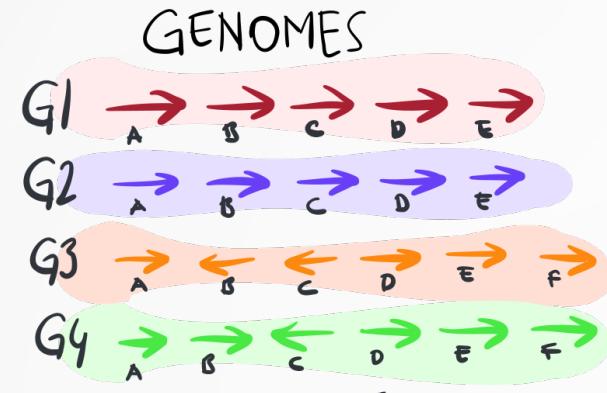


CALCULATING A PANGENOME





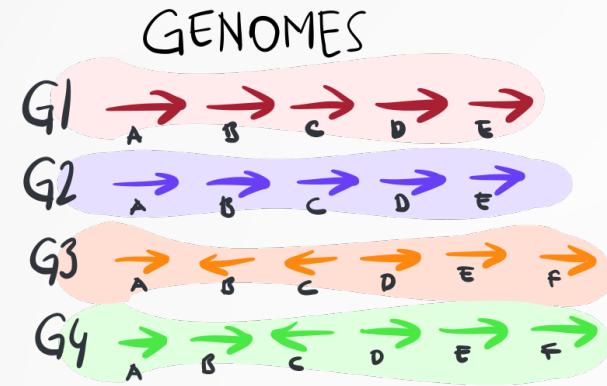
CALCULATING A PANGENOME



GENES FASTA



CALCULATING A PANGENOME

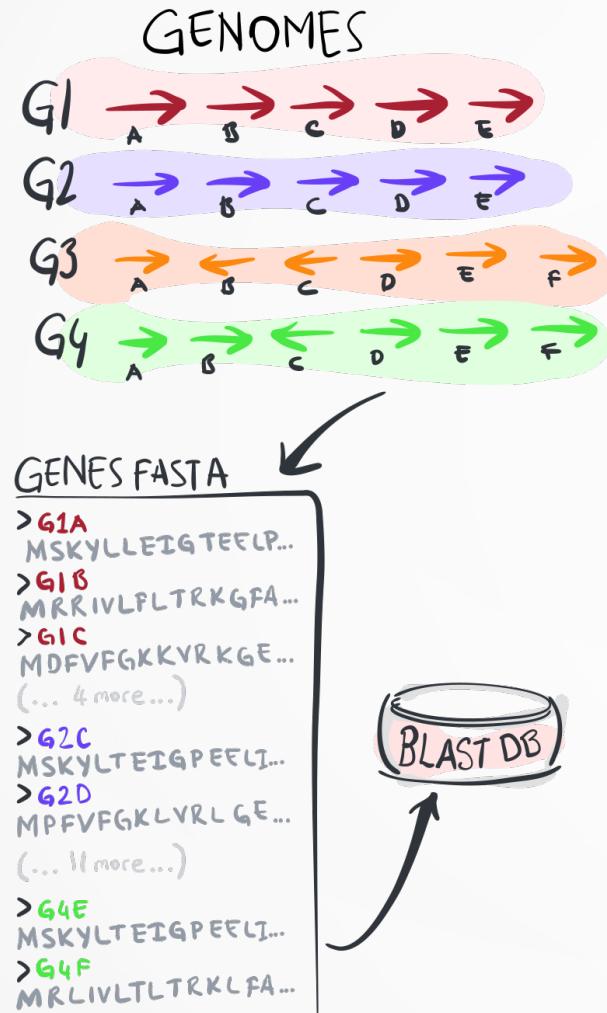


GENES FASTA

```
>G1A  
MSKYLLEIGTEELP...  
>G1B  
MRRIVLFLTRKGFA...  
>G1C  
MDFVFGKKVRKGE...  
(... 4 more...)  
>G2C  
MSKYLTEIGPEELI...  
>G2D  
MPFVFGKLVRLG...  
(... 11 more...)  
>G4E  
MSKYLTEIGPEELI...  
>G4F  
MRLIVLTLTRKLFA...
```

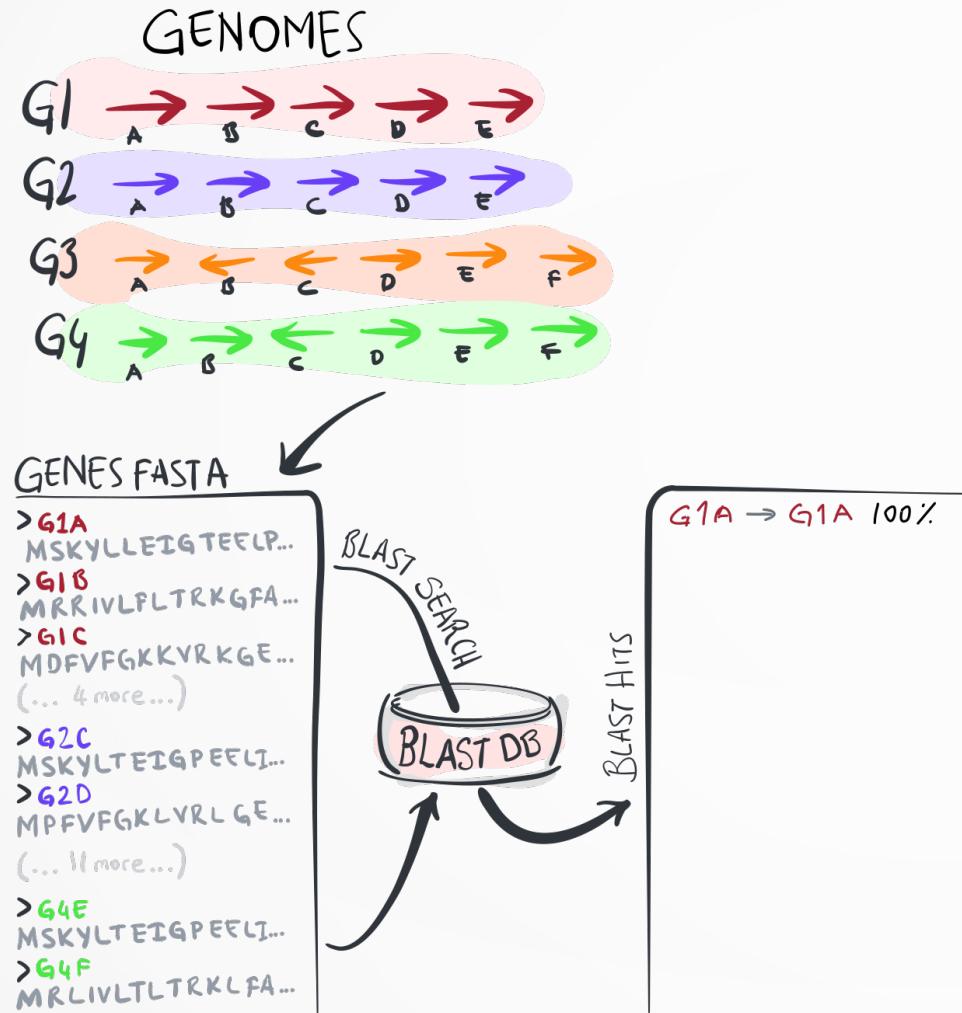


CALCULATING A PANGENOME



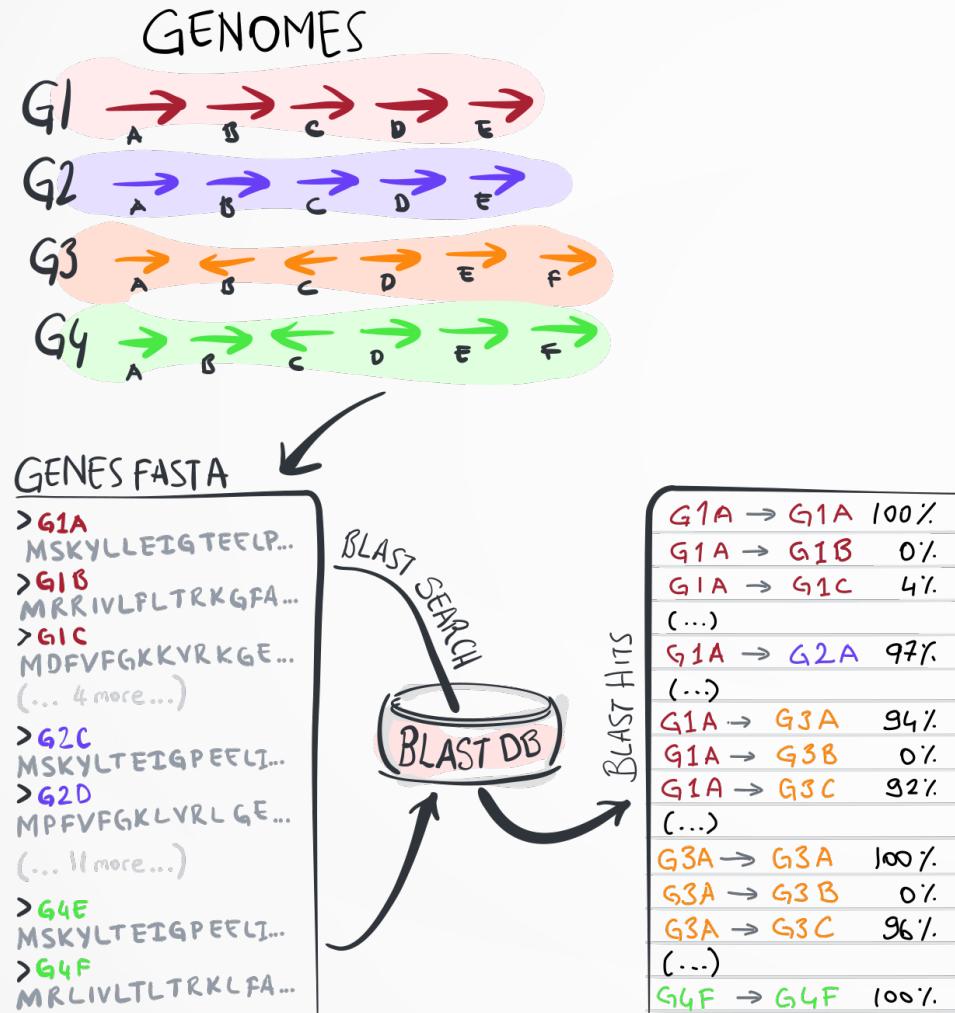


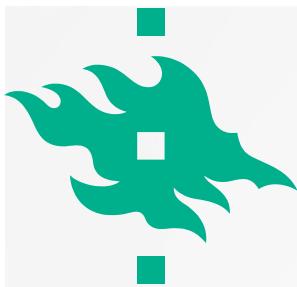
CALCULATING A PANGENOME



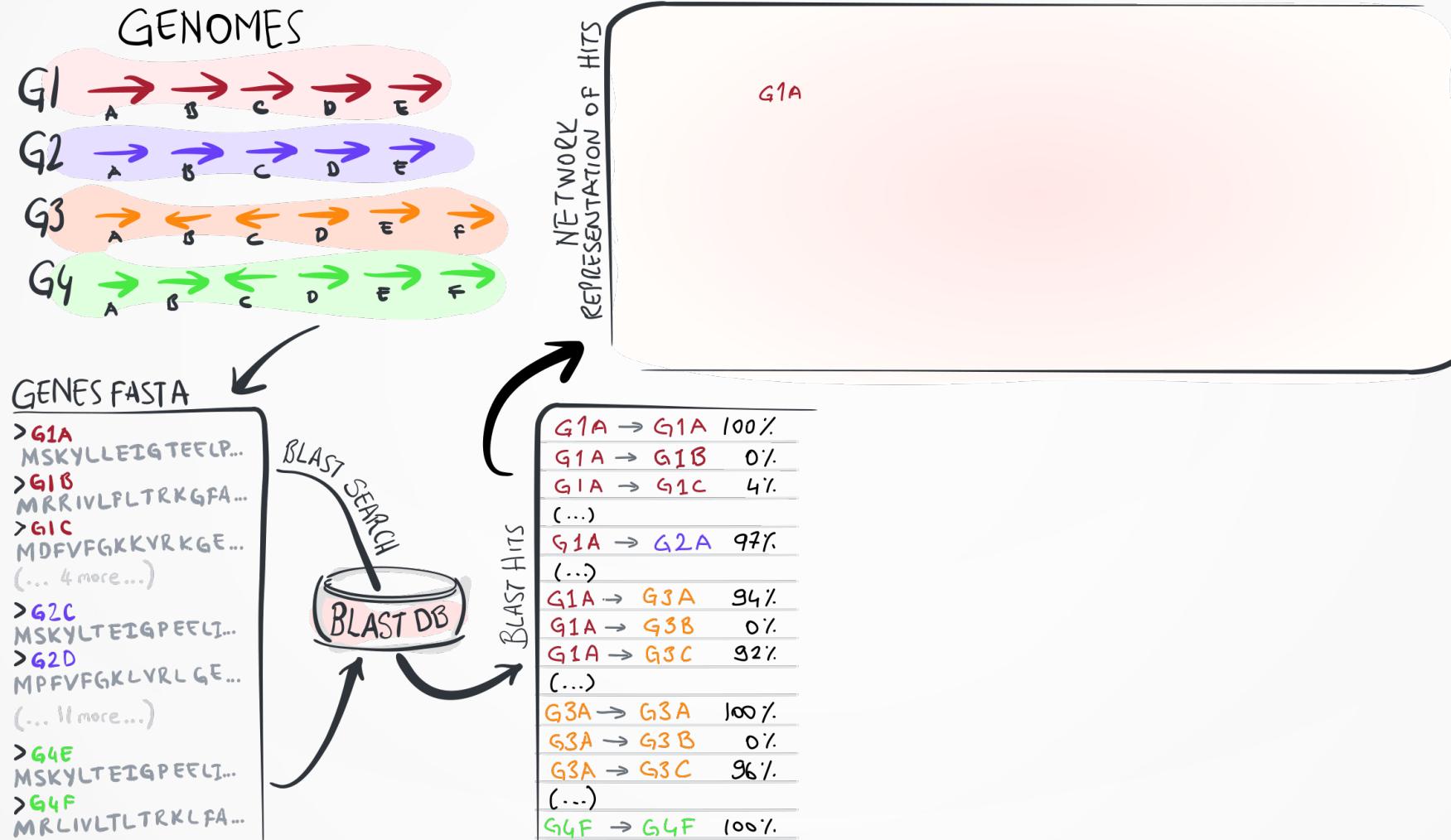


CALCULATING A PANGENOME



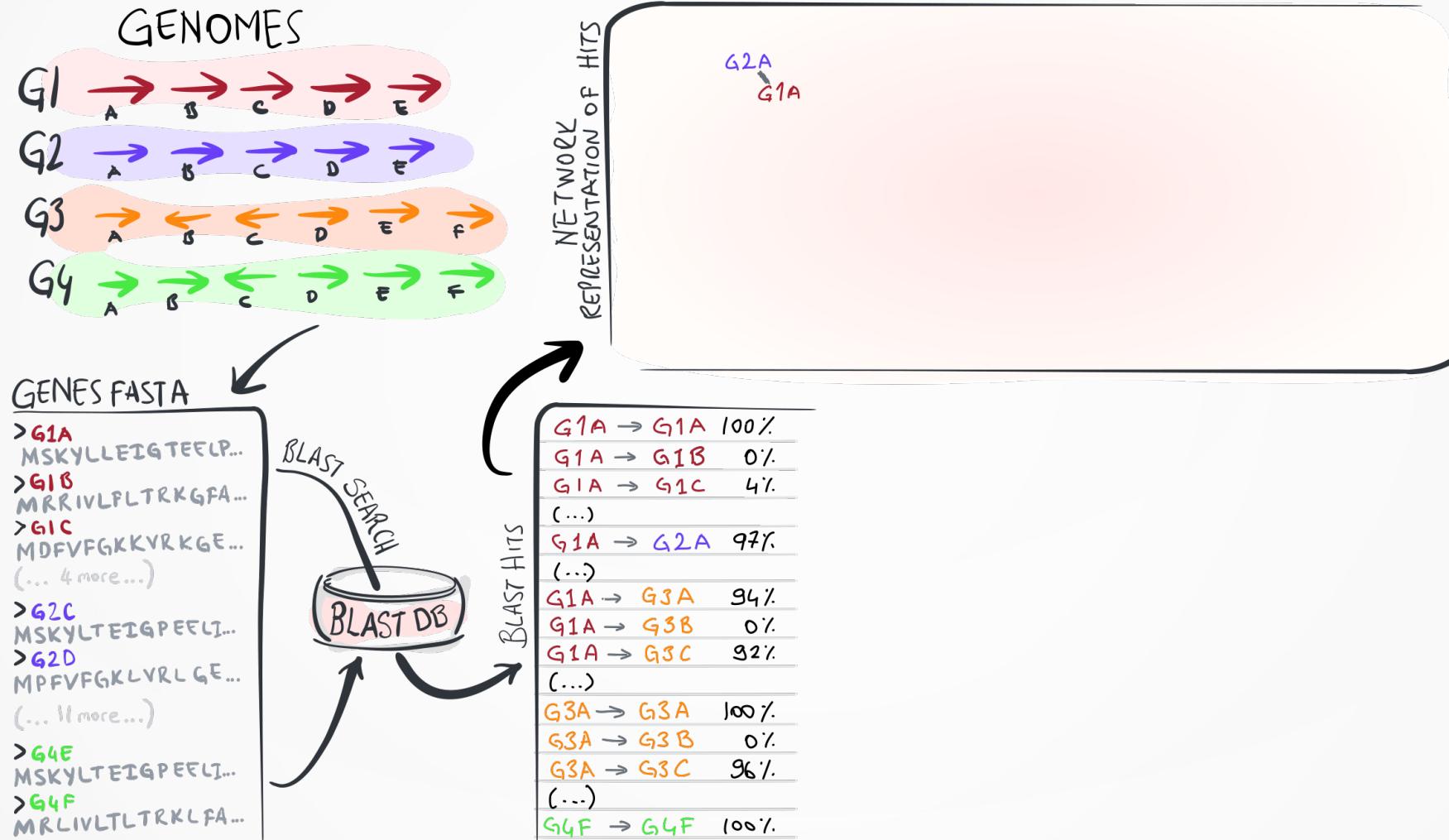


CALCULATING A PANGENOME



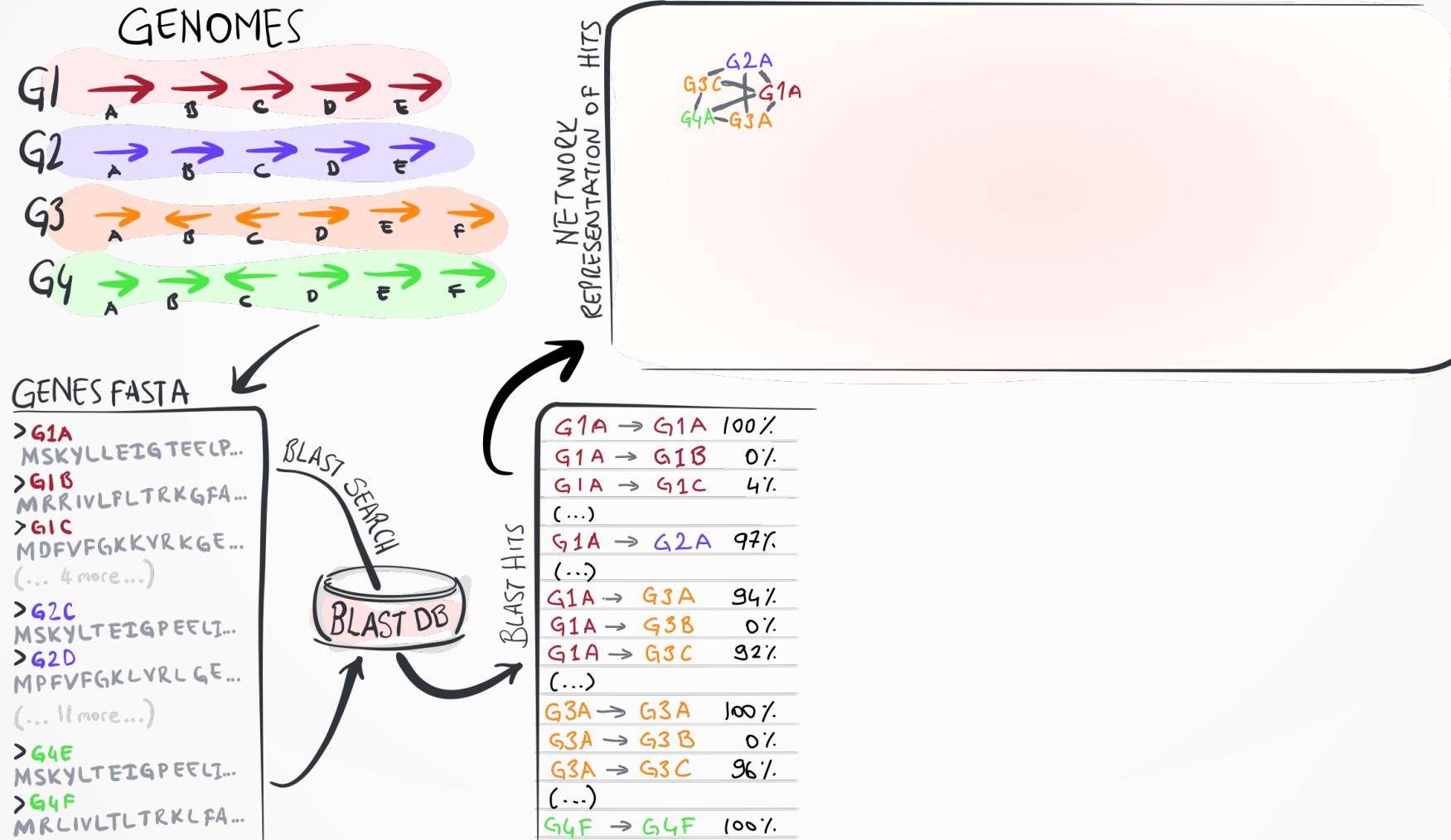


CALCULATING A PANGENOME



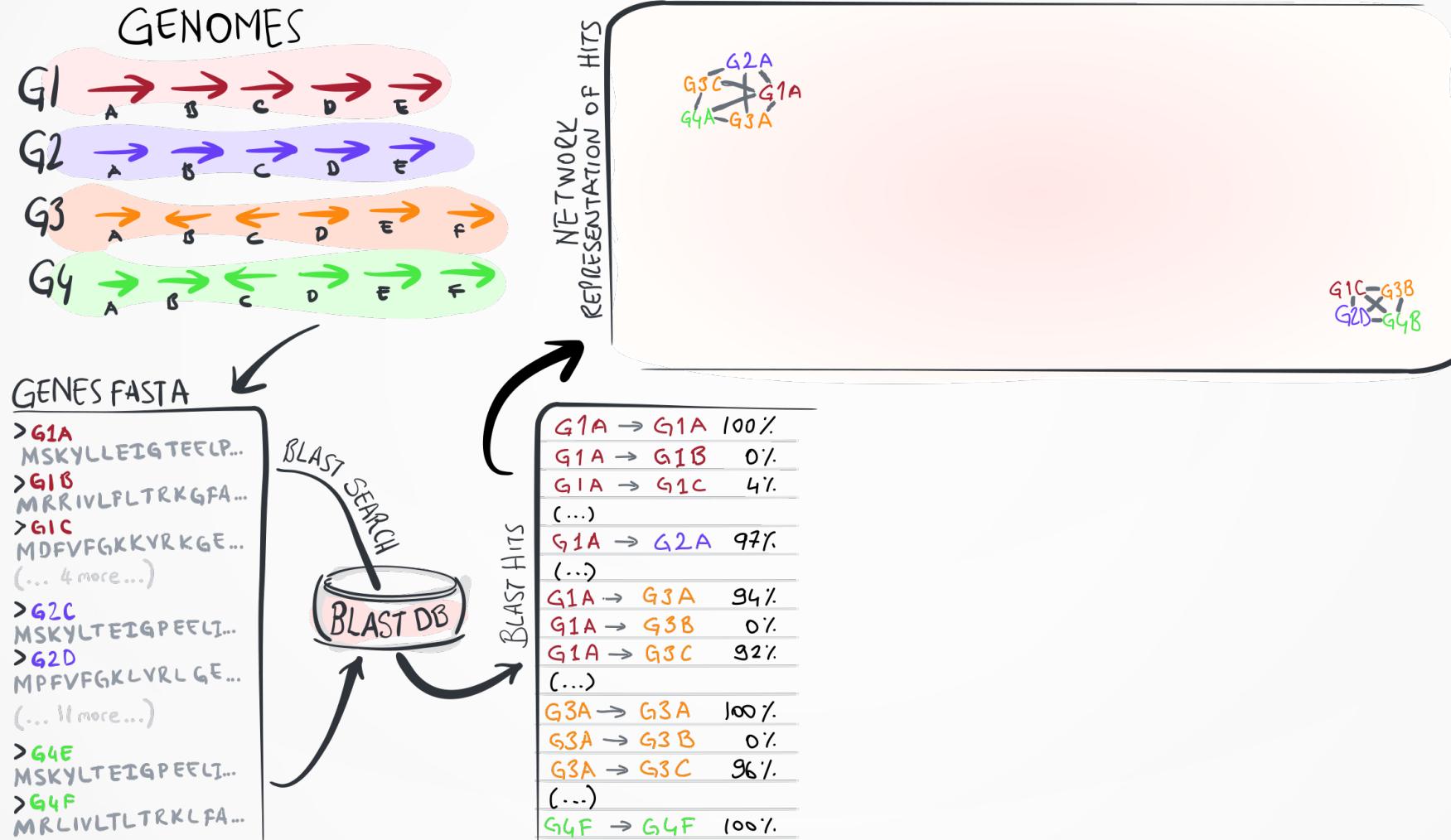


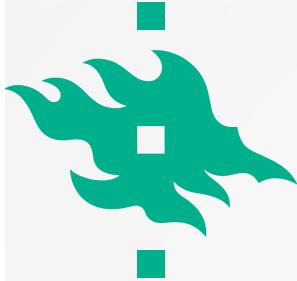
CALCULATING A PANGENOME



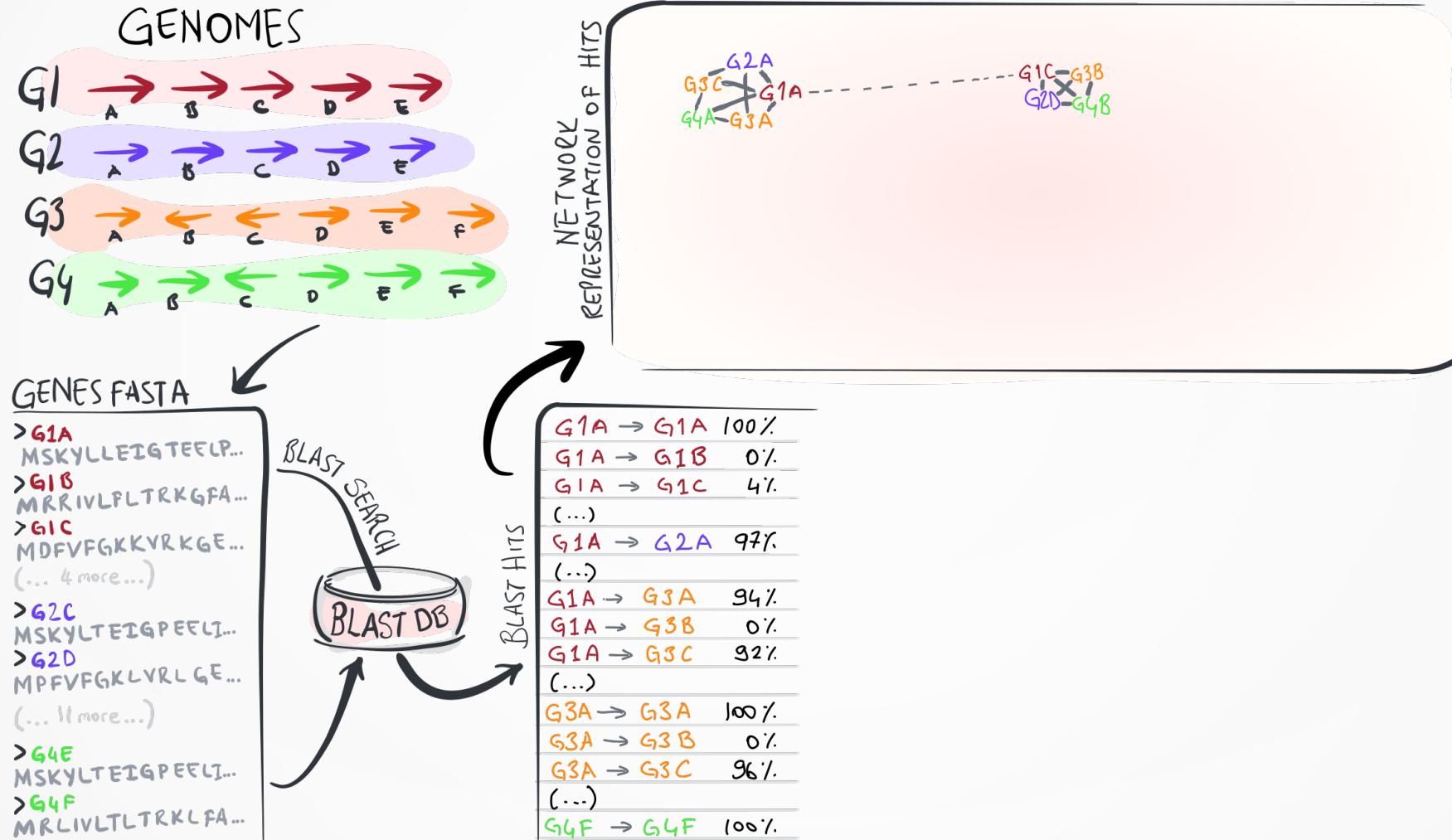


CALCULATING A PANGENOME



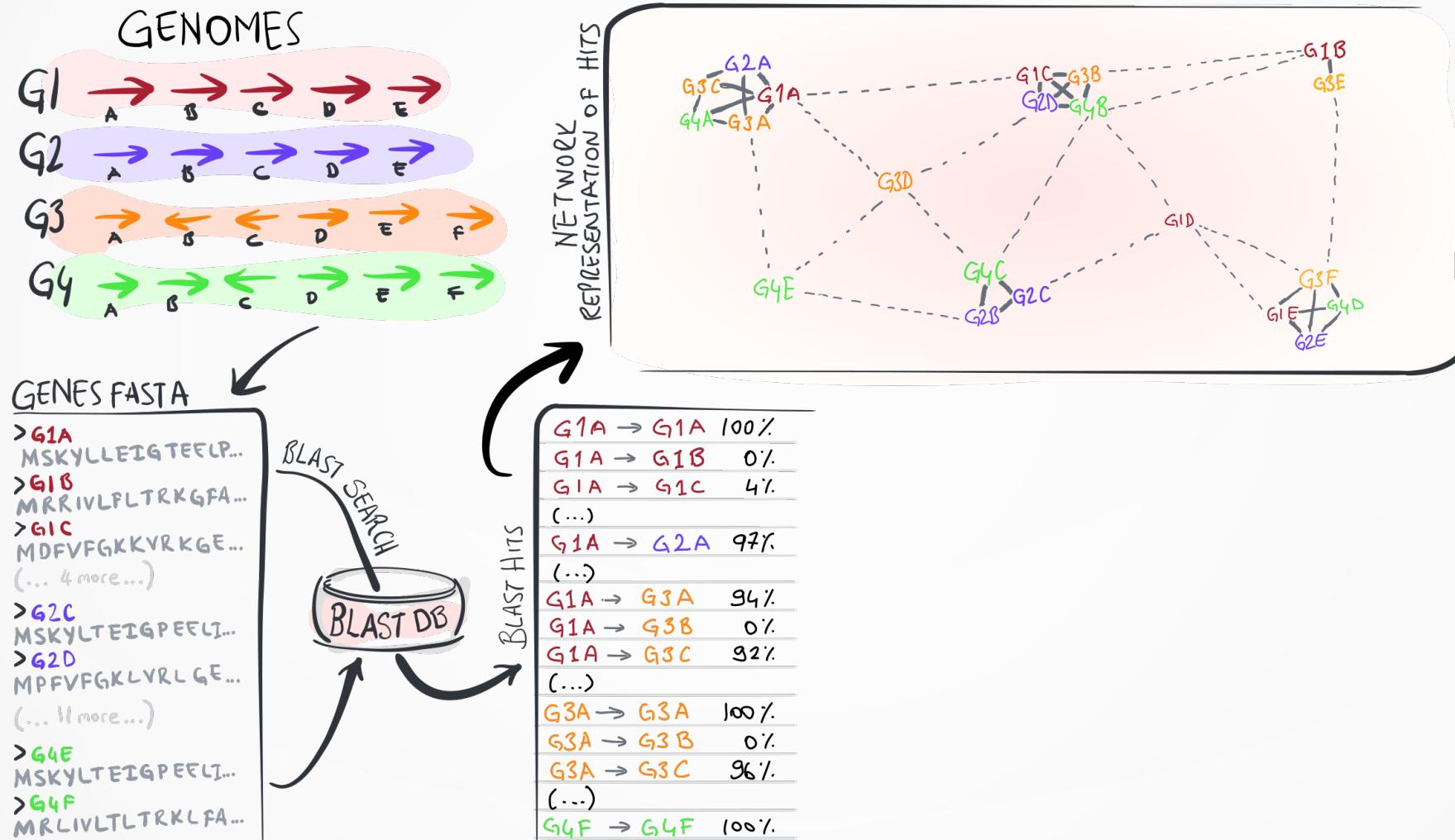


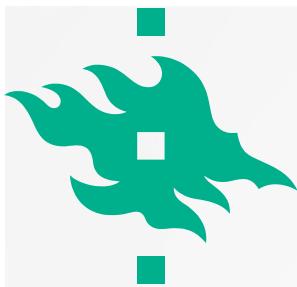
CALCULATING A PANGENOME



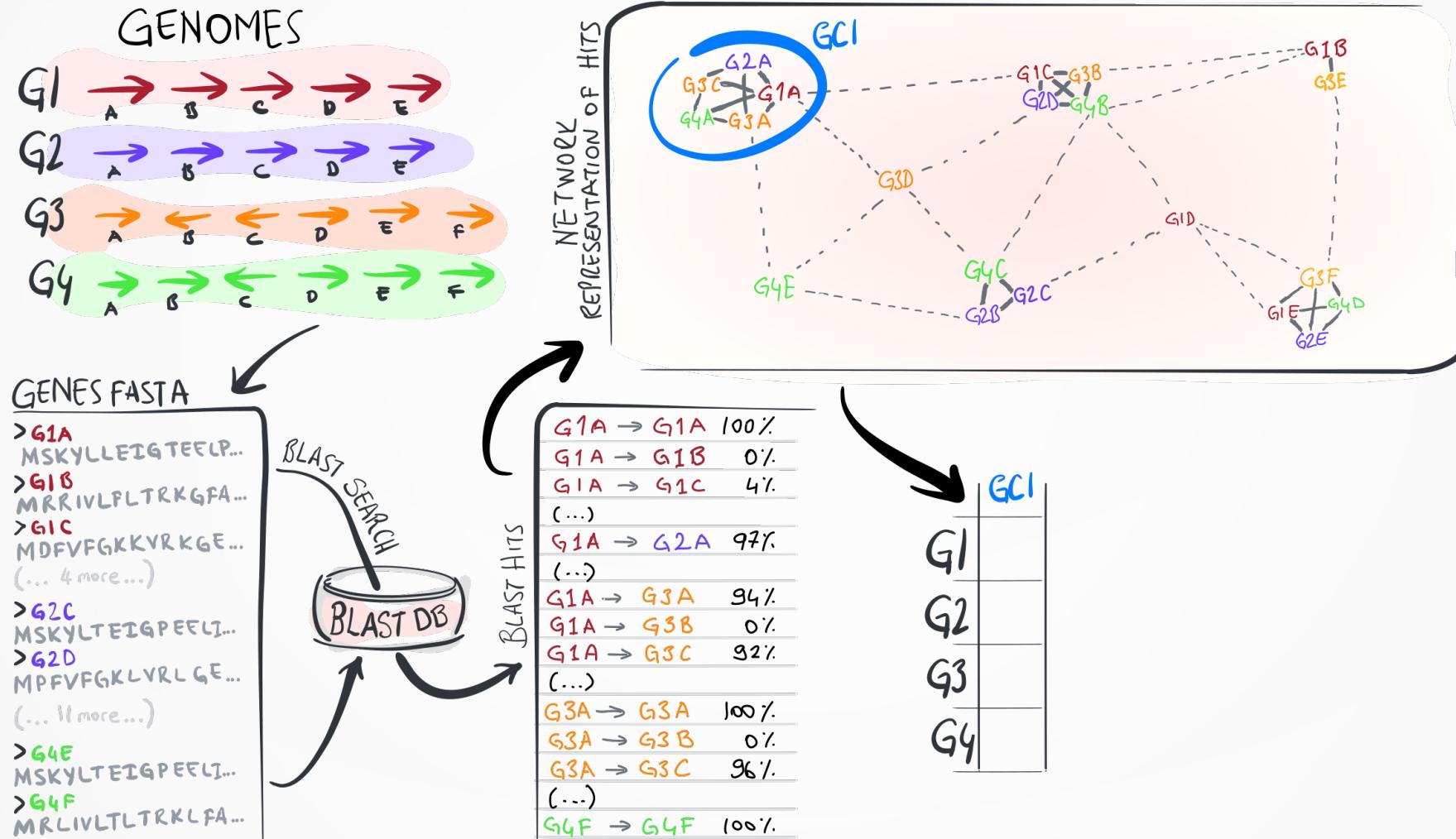


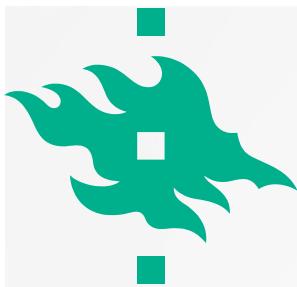
CALCULATING A PANGENOME



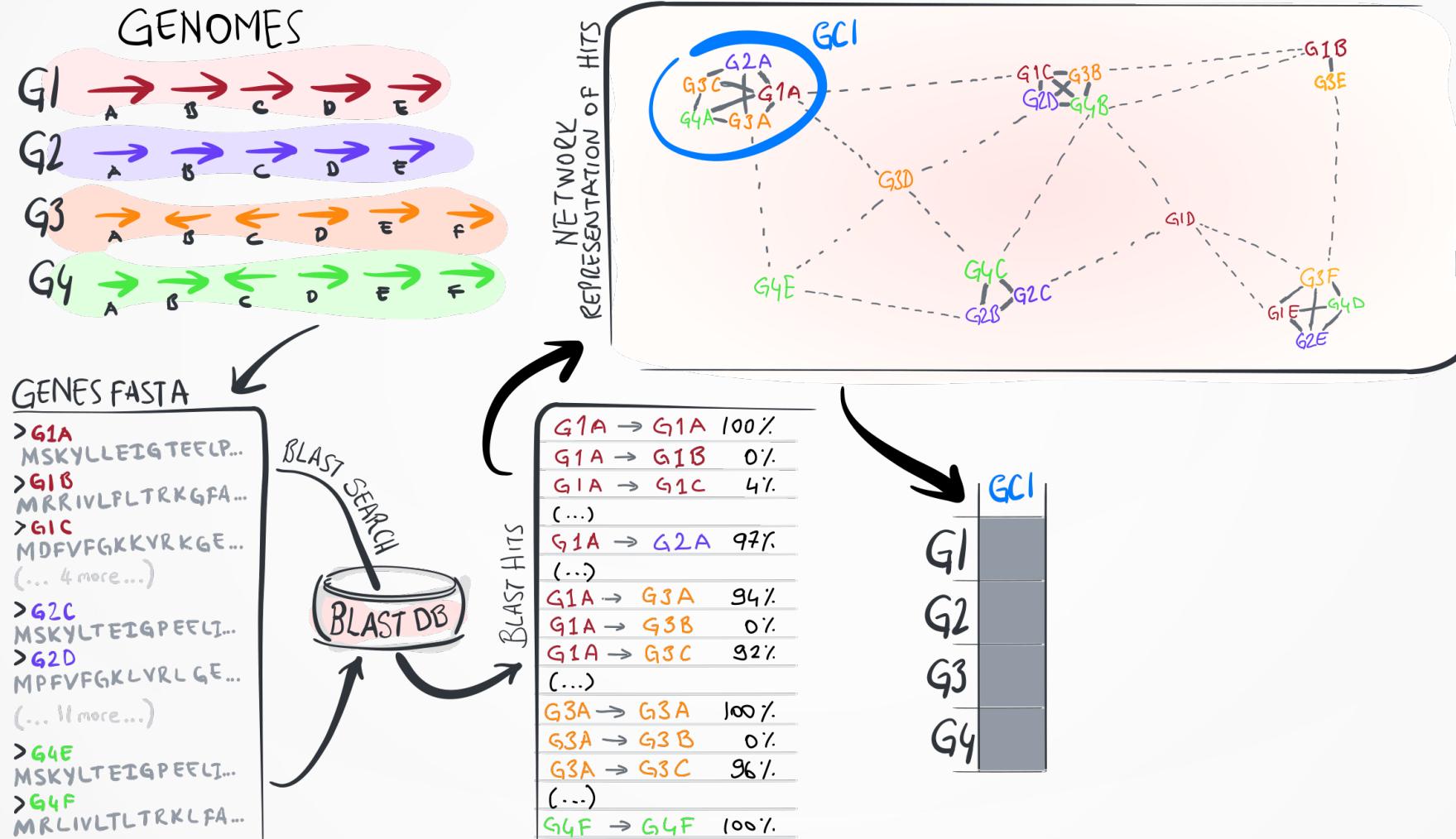


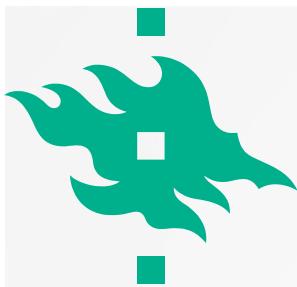
CALCULATING A PANGENOME



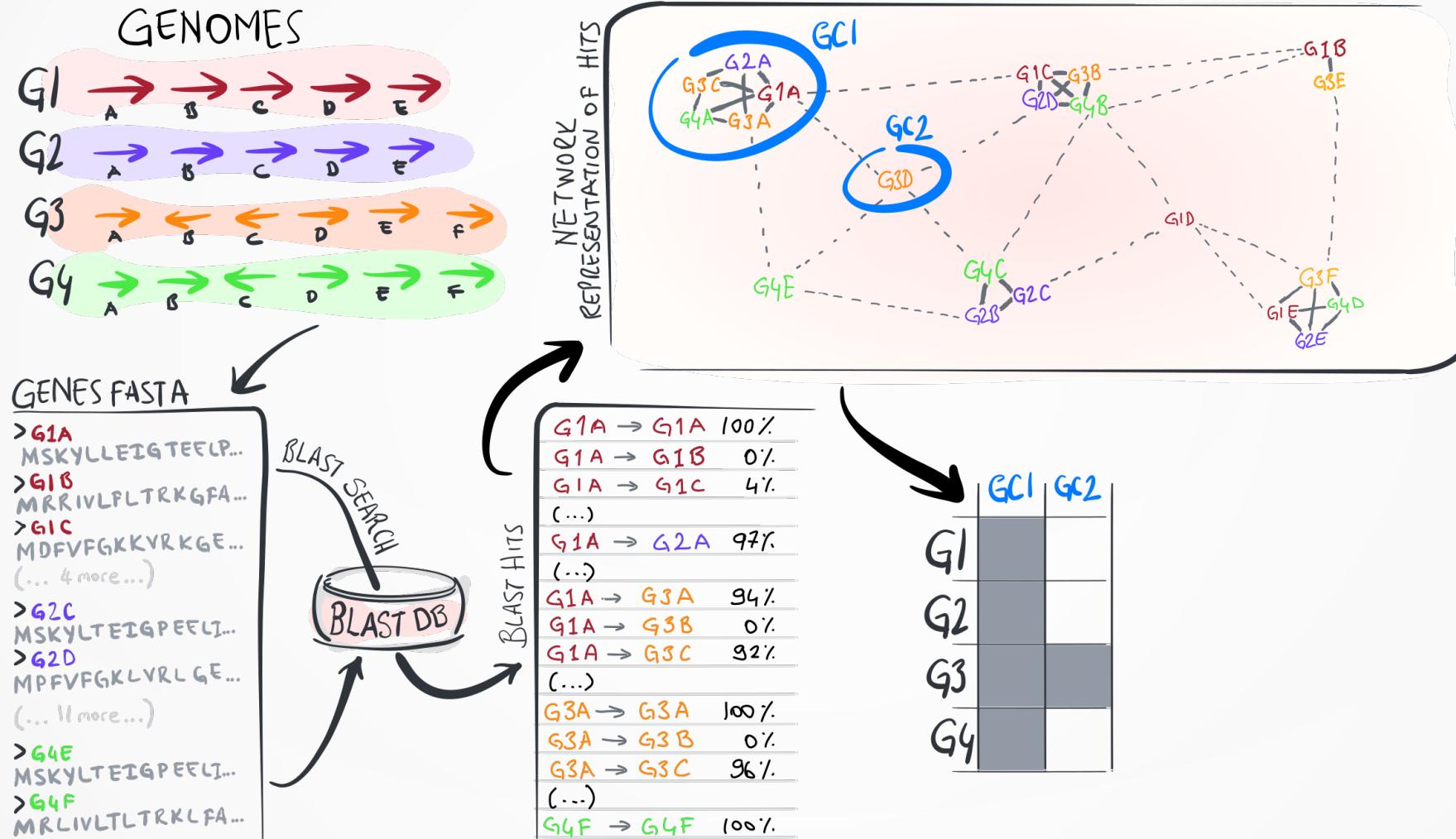


CALCULATING A PANGENOME



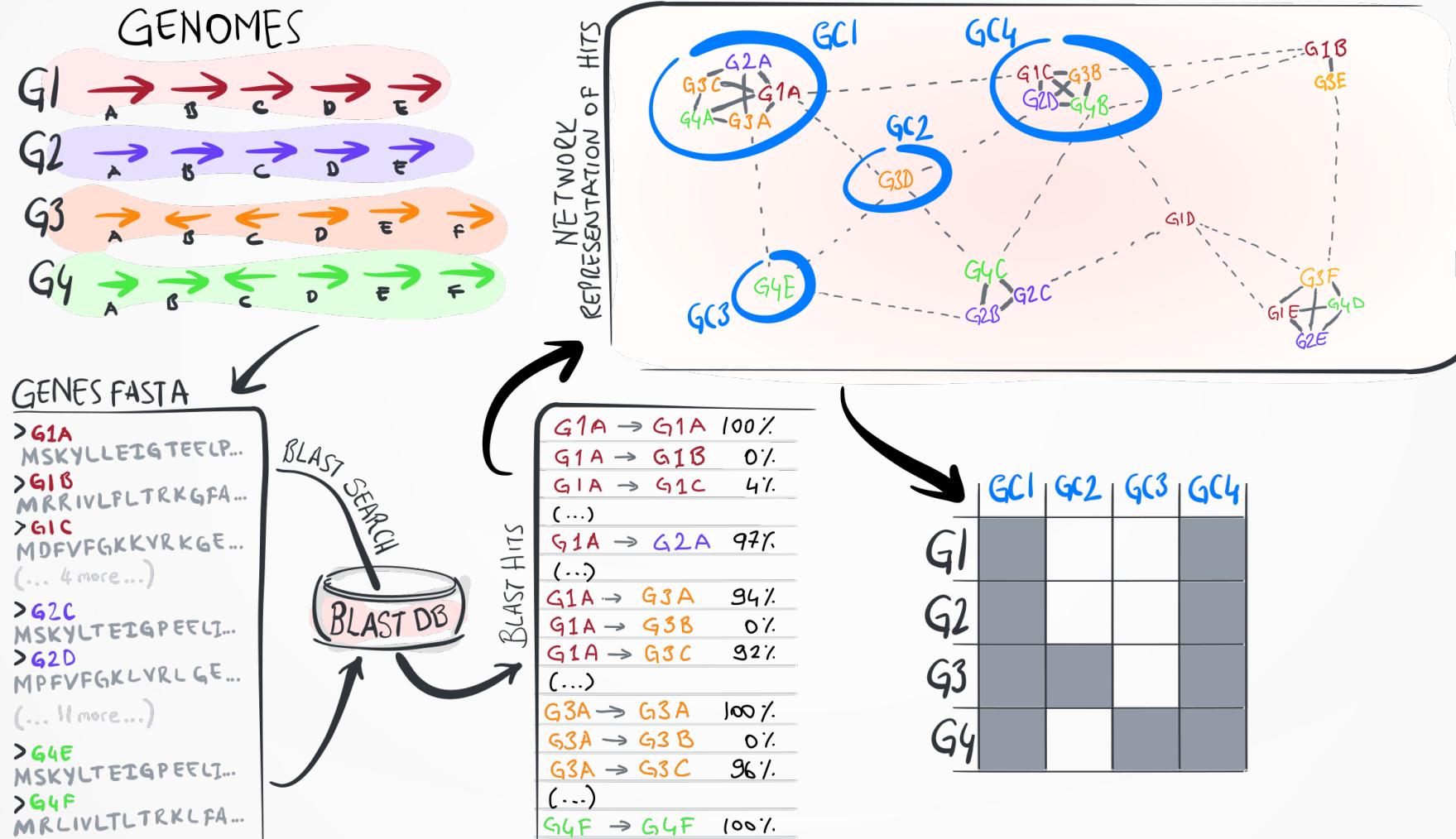


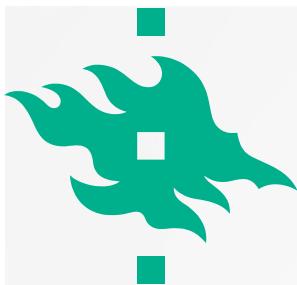
CALCULATING A PANGENOME



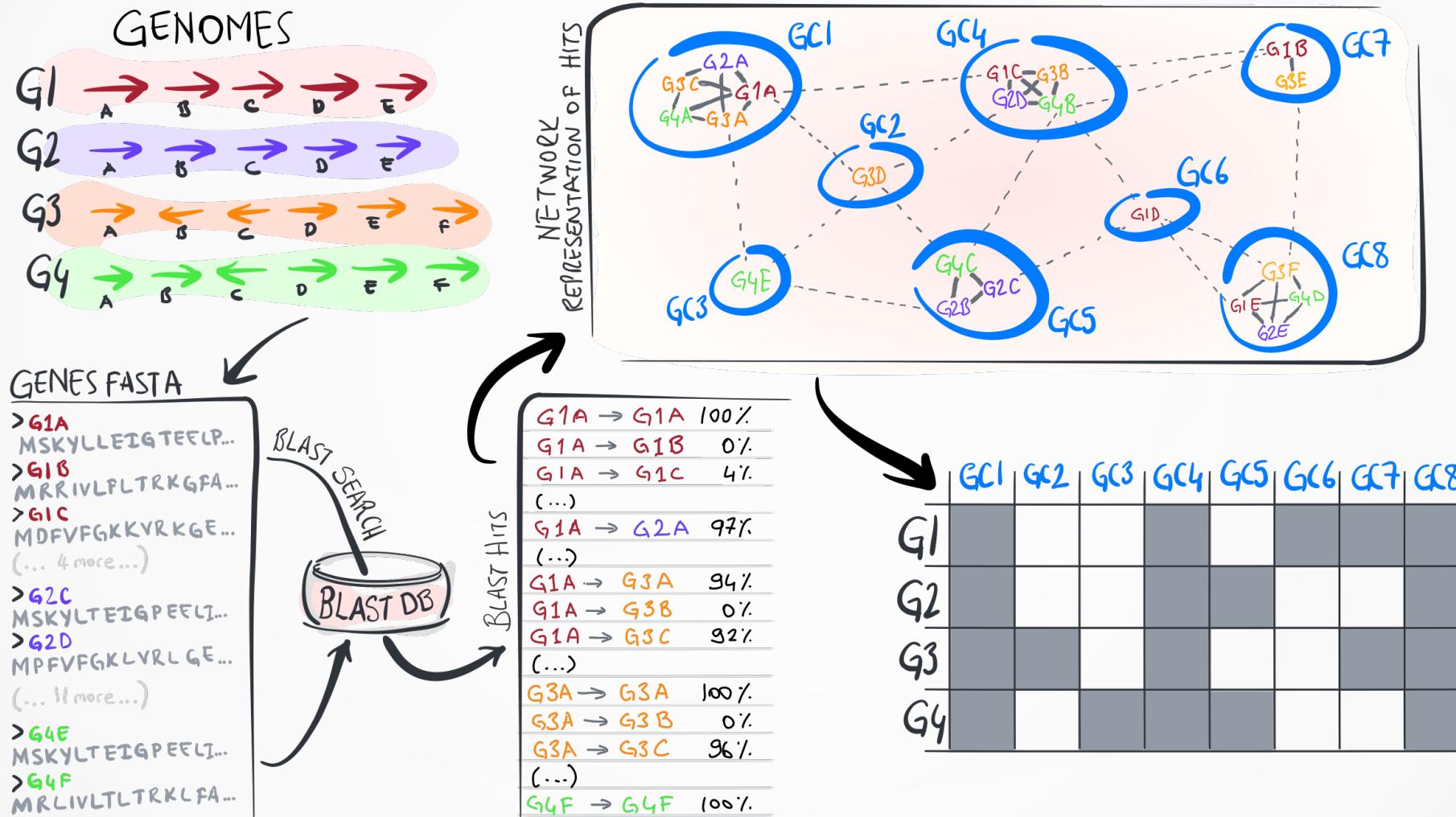


CALCULATING A PANGENOME



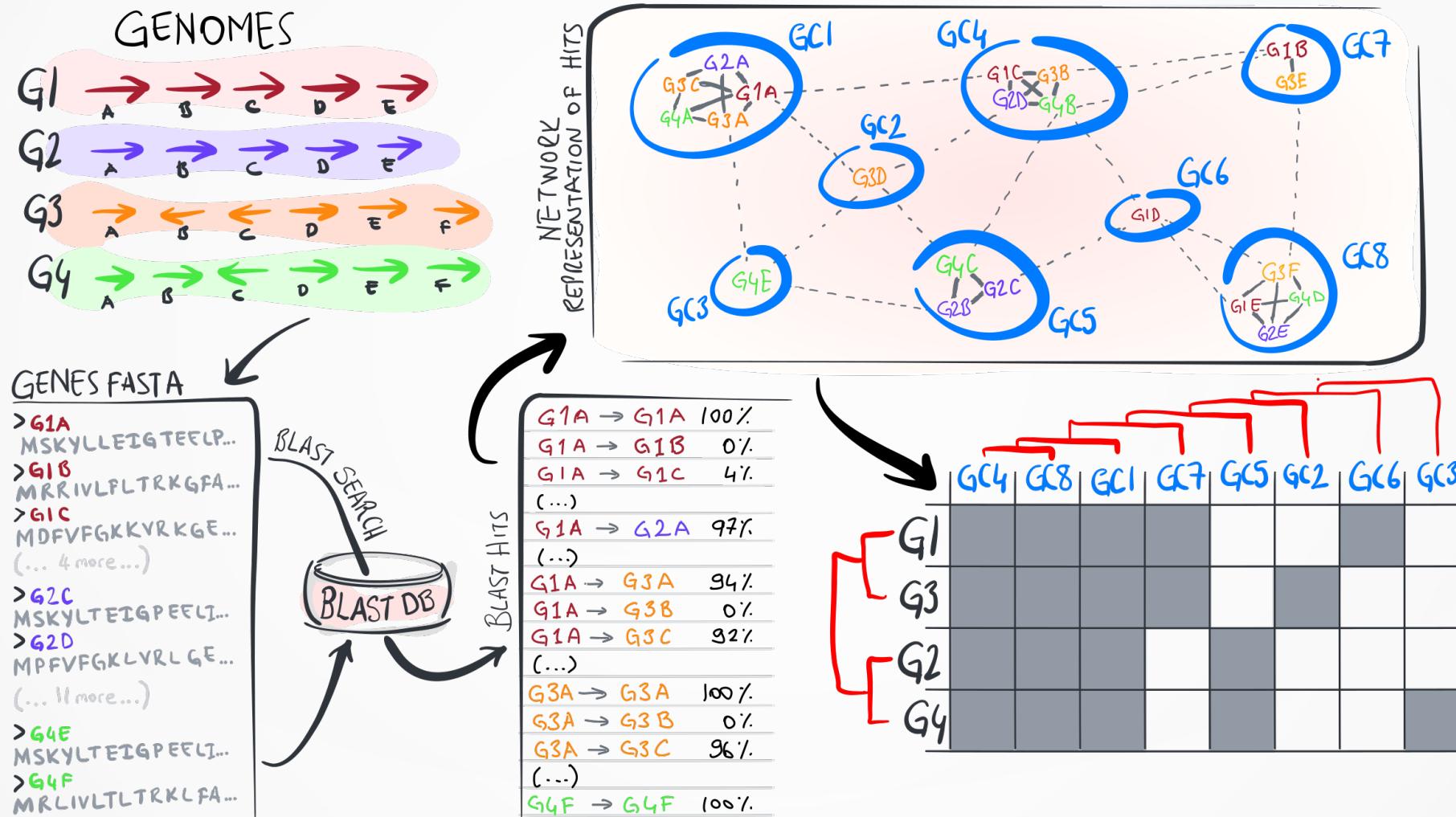


CALCULATING A PANGENOME

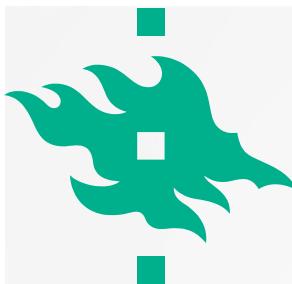




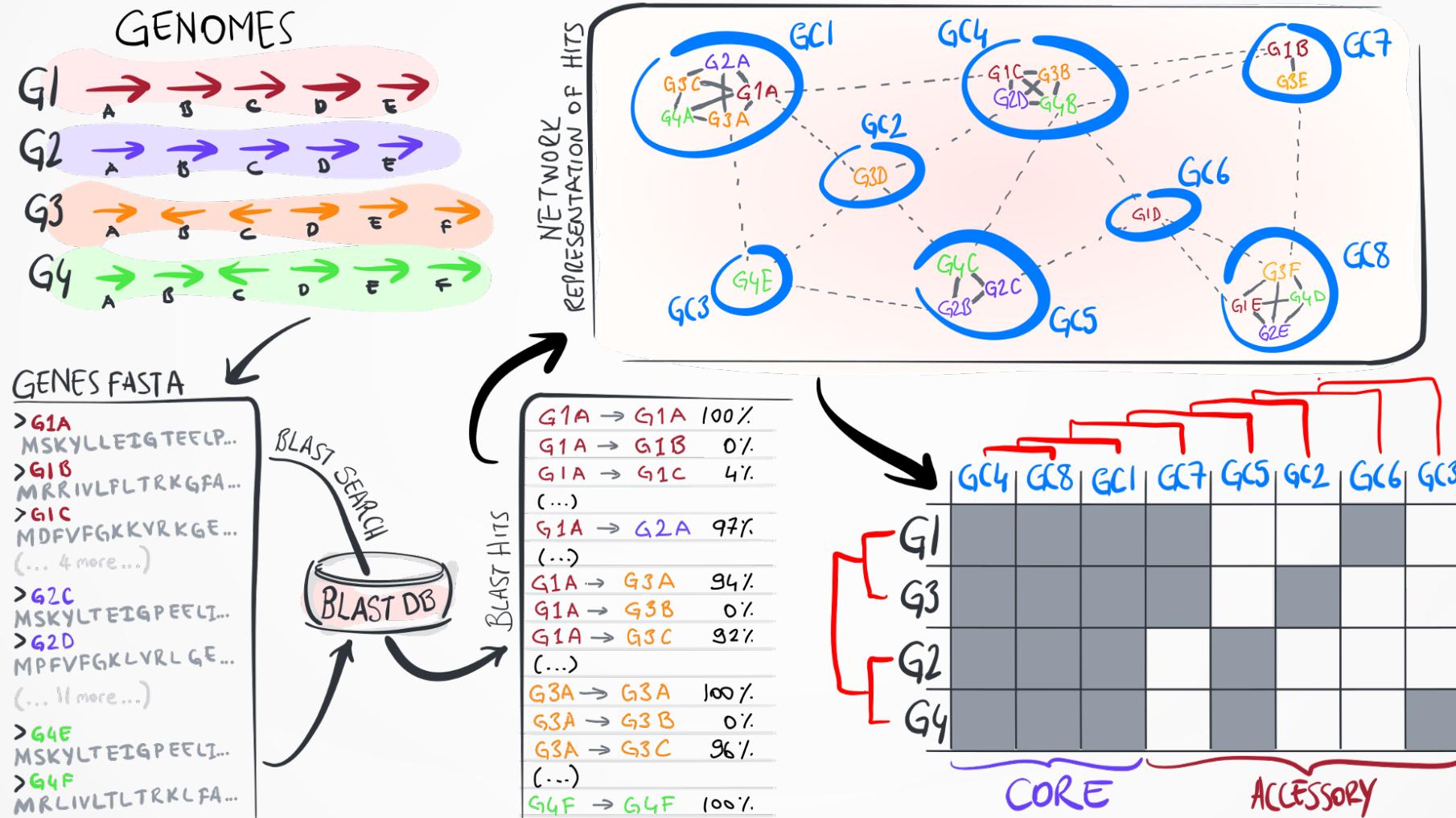
CALCULATING A PANGENOME

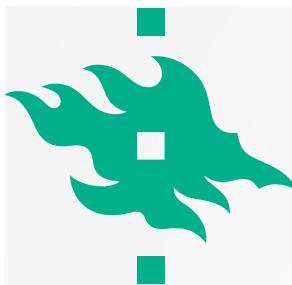


HELSINGIN YLIOPISTO HELSINGFORS UNIVERSITET UNIVERSITY OF HELSINKI



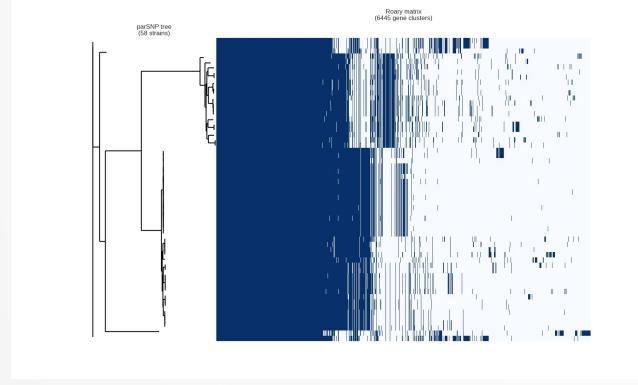
CALCULATING A PANGENOME





OTHER APPROACHES

Roary: the Pan Genome Pipeline



panX: Pan-genome Analysis & Exploration

Available pan-genomes
Selected species (see below for more) ▾
Search species

panX is a software package for comprehensive analysis, interactive visualization and dynamic exploration of bacterial pan-genomes. The analysis pipeline is based on DIAMOND, MCL and phylogeny-aware post-processing.

The visualization application encompasses various interconnected components (statistical charts, gene cluster table, alignment, comparative phylogenies, metadata). Gene clusters can be rapidly searched and filtered by summary statistics such as annotation and phylogenetic properties such as diversity.

Metadata and gene presence/absence patterns can be mapped onto the species tree. Such mapping facilitates the identification of genes associated with phenotypes such as antibiotic resistance, virulence, or epidemiological parameters such as host age.

Showcases of panX functionalities:

Epidemiology of 616 Streptococcus pneumoniae strains
Croucher et al. (2015) Demo Explore

panX - pan-genome analysis and exploration from panX

Diverse Prochlorococcus genomes: gene gain/loss pattern Bitter et al. (2014) Explore

EUPAN, micropan, NGSPanPipe, PanACEA, Panaconda, PanCake, PanFunPro, PanGeT, PanGFR-HM, PanGP, PANINI, PANNOTATOR, PanOCT, Panseq, Pan-Tetris, PanTools, PanViz, PanWeb, seq-seq-pan, ...