
Catherine Burke
UTS


Michael Liu
UTS


Chris Beitel
UTS, UC Davis


Matt DeMaere
UTS

Also:
Sunaina
Melissa Gardiner
UTS

Metagenomic 3C, full length 16S amplicon sequencing on Illumina, and the diabetic skin microbiome

Presented by: Catherine Burke and Aaron Darling
twitter: @koadman

ithreeinstitute
infection·immunity·innovation

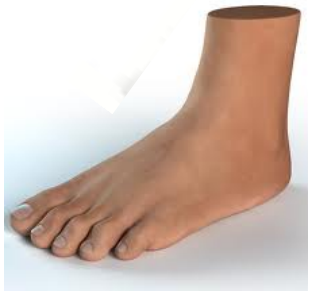
UNIVERSITY OF TECHNOLOGY SYDNEY

UCDAVIS
UNIVERSITY OF CALIFORNIA

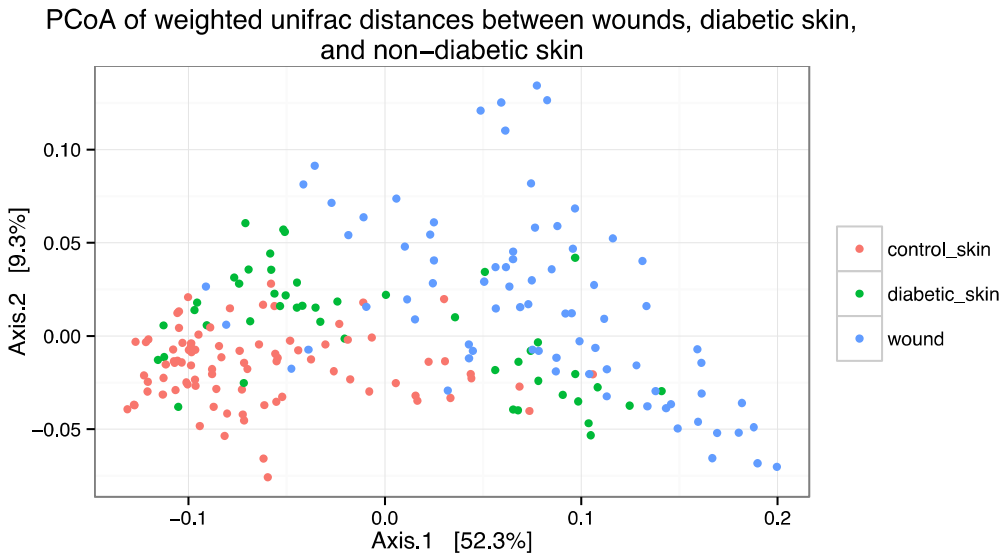
Skin and wound microbiome in type II diabetes

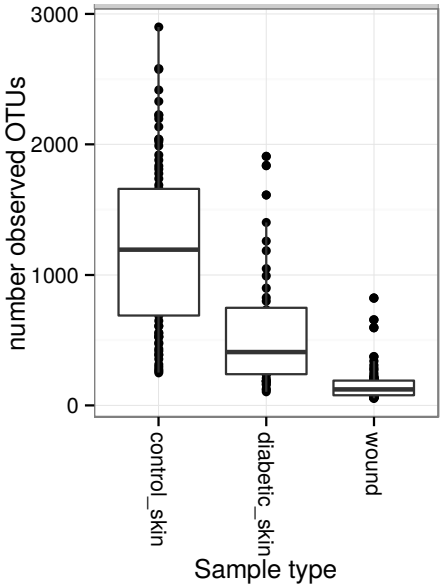
- Does diabetes affect the composition of the skin microbiome, and does this affect microbial colonisation of chronic wounds?

- 10 diabetic and 10 control subjects
  - Sampled every 2 weeks over a 12 week period
    - Skin (foot), wounds (diabetics)



Skin and wound microbiome in type II diabetes





- Main difference between control and diabetic skin is decreased diversity (at this resolution…..)

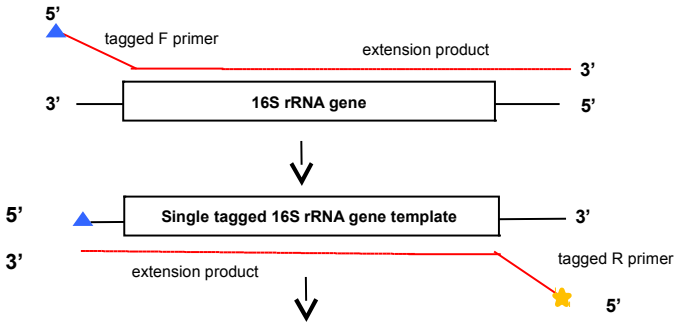
High throughput long 16S sequencing on Illumina MiSeq

- Sequencing of full length 16S sequences via addition of unique molecular tags, fragmentation, sequencing and reconstruction.
  - Higher quality sequences
  - Better phylogenetic resolution
  - Removal of PCR artifacts
  - Improvement on PCR bias

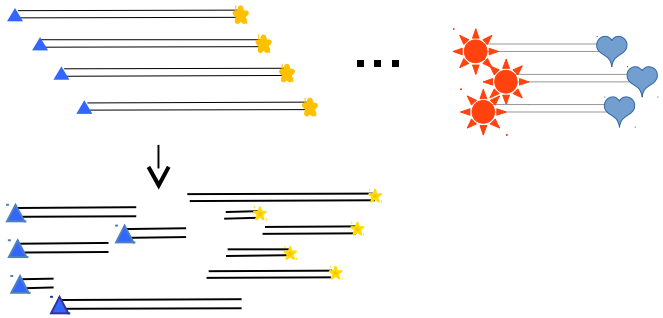
- Removal of PCR artifacts and amplification bias

High throughput long 16S sequencing on Illumina MiSeq

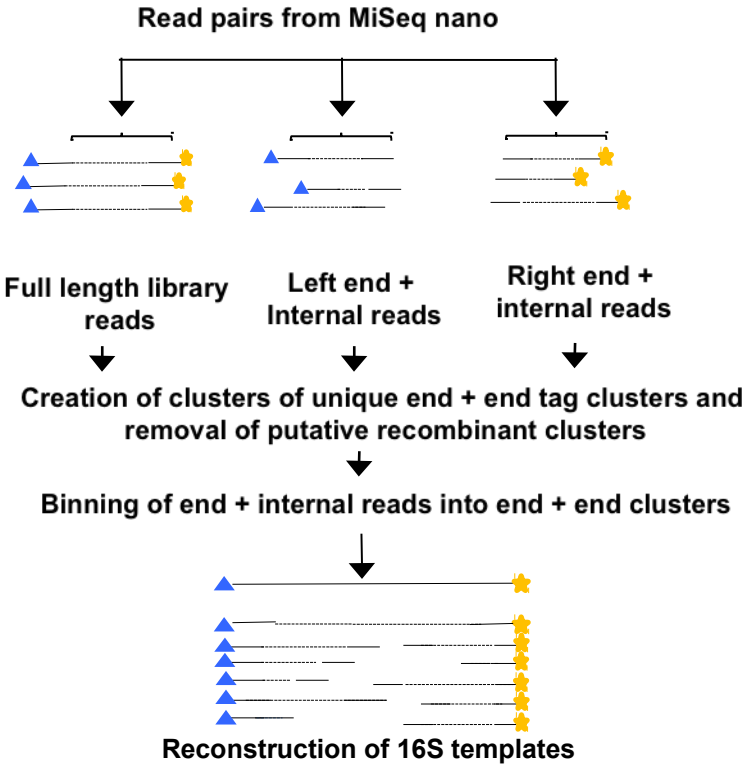
Tagging with long primers



Amplification and fragmentation of tagged templates



Sequencing and assembly



Is metagenomics the answer...??

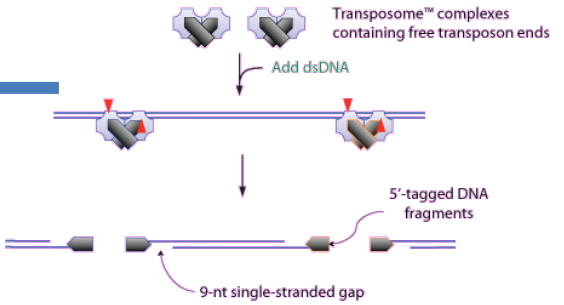
Sample

Smash with bead beater

Purify DNA

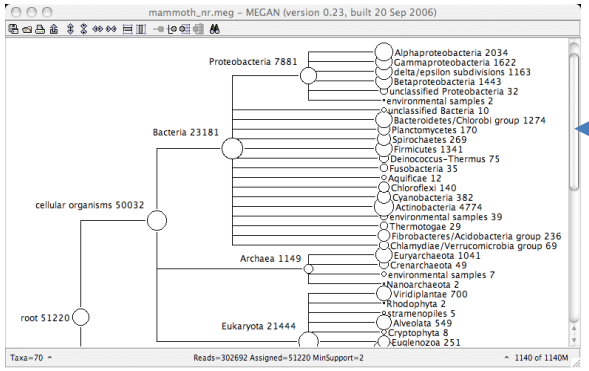
Shear DNA to 400nt frags

Prep sequencer library



Sequence (yay!)

Analyze



Is metagenomics the answer...??

Sample

Smash with bead beater

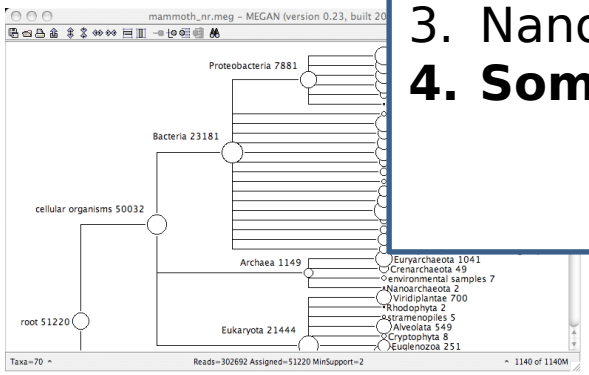
Purify DNA

Shear DNA to 400nt frags

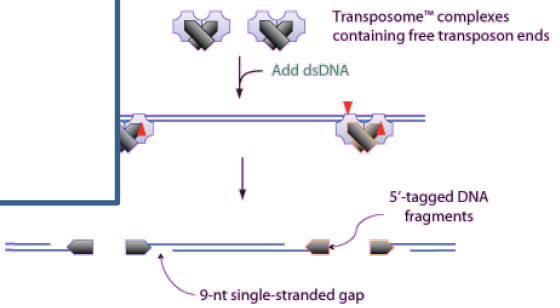
<u>At least three paths forward:</u>

1. Physically “dissect” microbial communities (e.g. Rinke *et al* 2013)
2. Better inference methods
3. Nanopore long read sequencing?
4. **Something magical?**

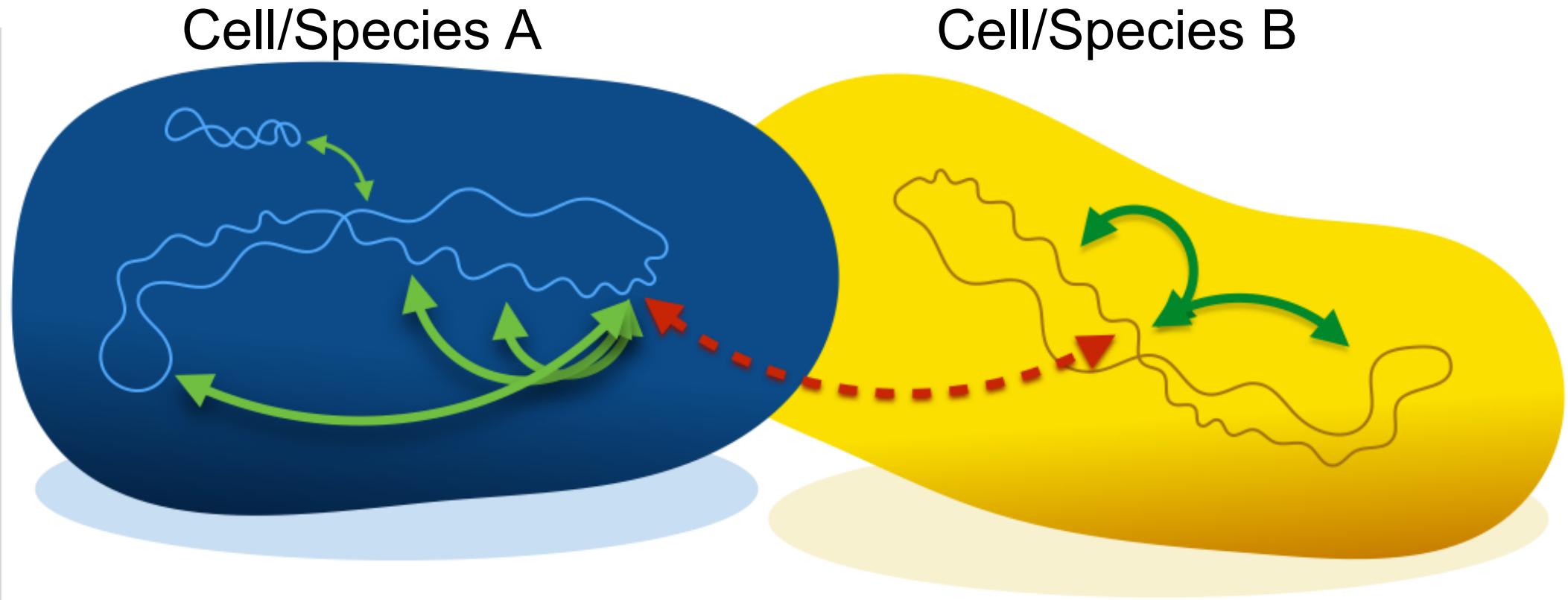
Analyze



sequencer library



We propose to use 3C / Hi-C for metagenomics

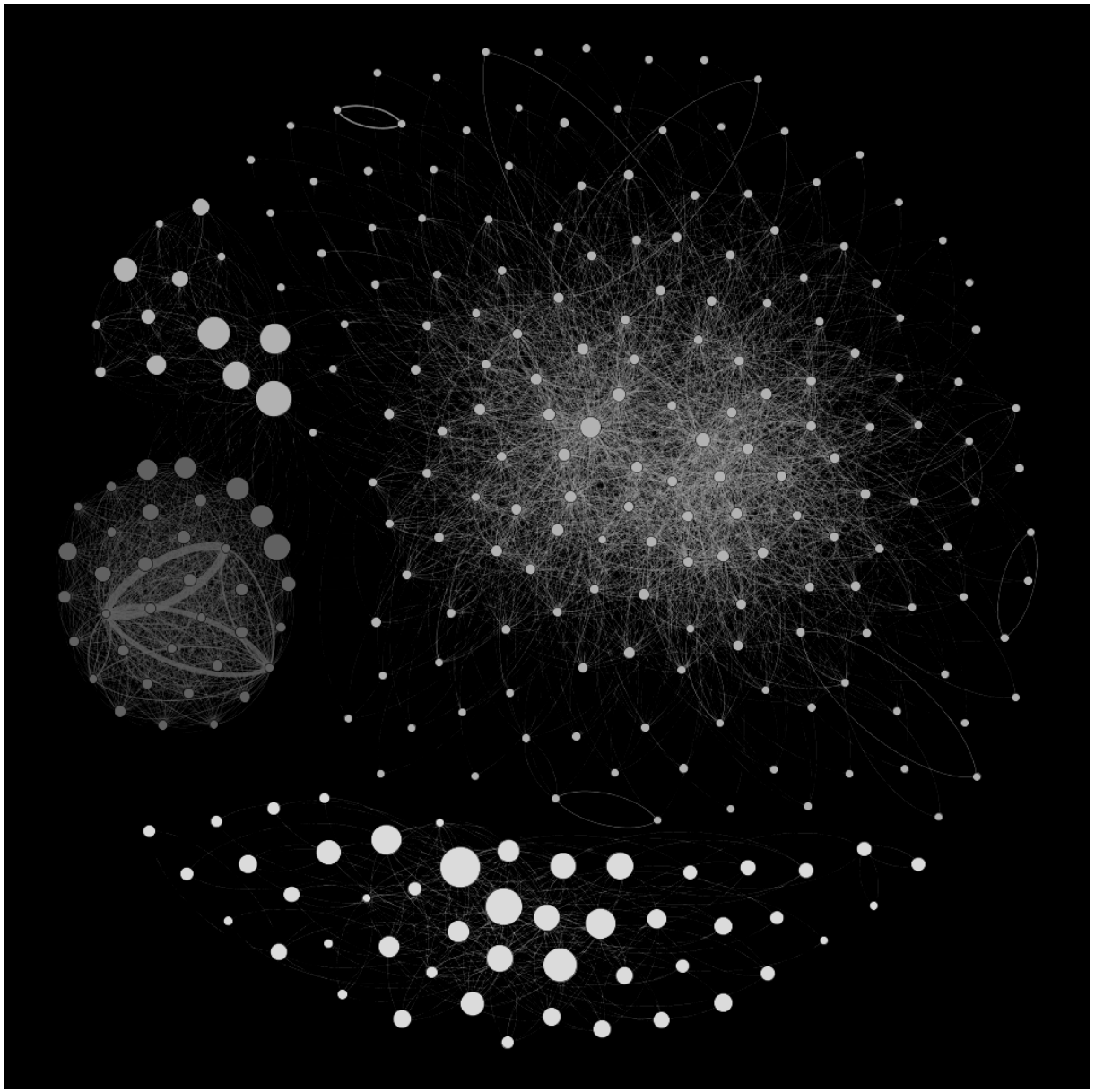


Conjecture: DNA in the same cell will become crosslinked, ligated, and sequenced more frequently (green links) than DNA fragments from different cells (red links)

A little test: Mix up isolate cultures of four species (five strains) with finished genomes, apply Hi-C, sequence on MiSeq. Can reconstruct the input genomes

A Hi-C scaffold graph

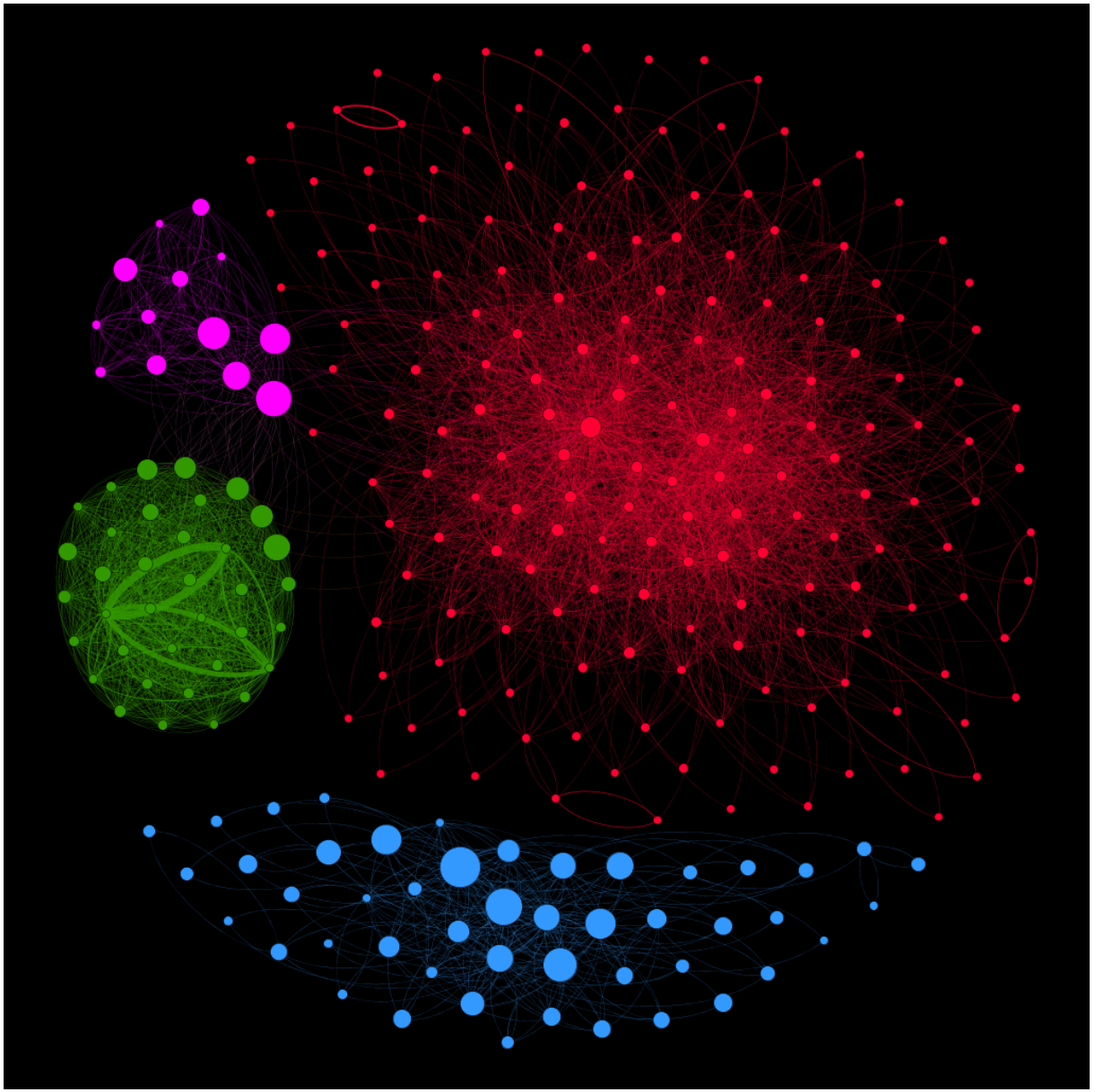
- 557 assembly scaffolds
- Nodes are scaffolds
- node size ∝ scaffold size
- edge weights ∝ normalized Hi-C read counts linking scaffolds
- Fruchterman-Reingold layout



Beitel *et al* 2014 *PeerJ*.

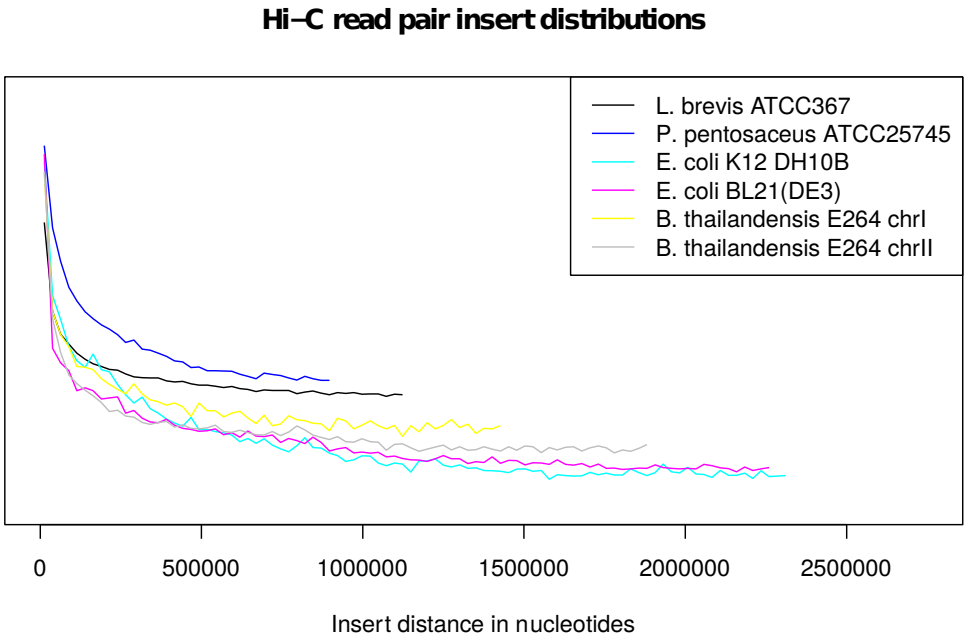
A Hi-C scaffold graph

- 557 assembly scaffolds
- Nodes are assembly contigs
- node size ∝ contig size
- edge weights ∝ normalized Hi-C read counts linking contigs
- Fruchterman Reingold layout
- Nodes colored by SPECIES
- Can we actually compute clusters?
- Markov clustering, I=1.1: 4 clusters, >97% of genome in each bin

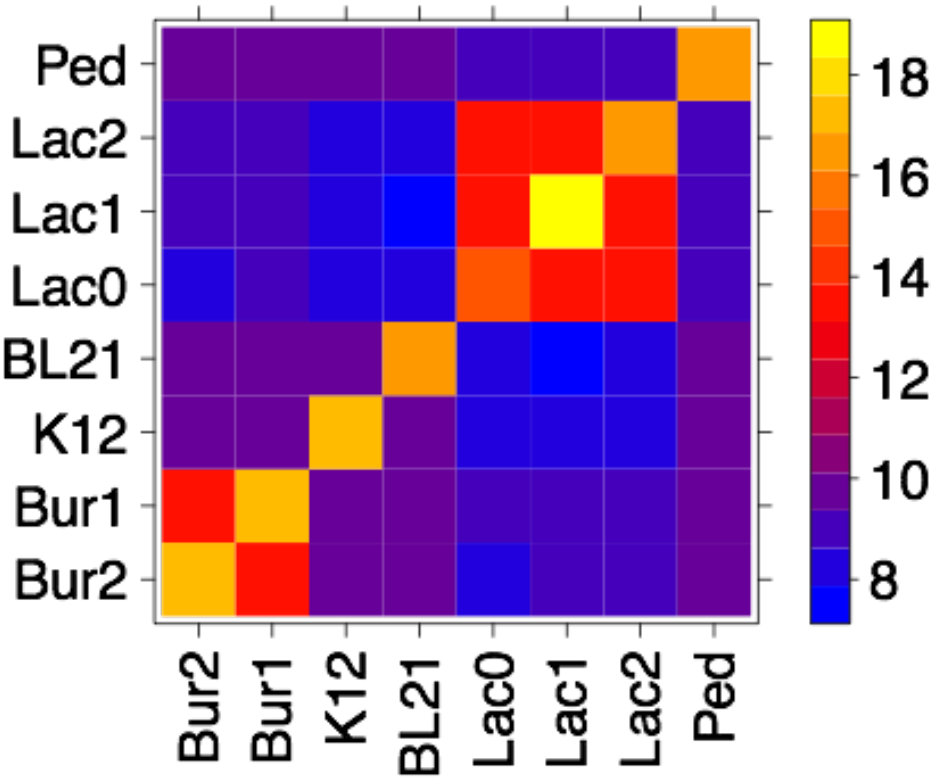


Beitel *et al* 2014 *PeerJ*.

Contig clustering: why it works



Rate at which Hi-C reads associate within and across species



- 99% of Hi-C read pairs associate within-species
- Hi-C read pairs associate throughout the chromosome
- Spans distances that no existing or imaginary “long read” technology can cover

Beitel *et al* 2014 *PeerJ*.

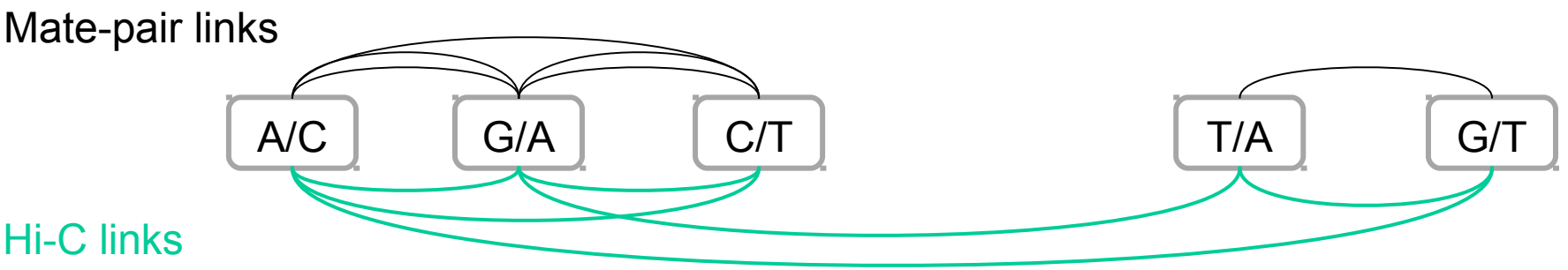
Resolving strain differences: Single nucleotide variants

Define a variant graph:

- Sites containing SNPs between *E. coli* K12 and BL21 are nodes
- Edges link SNP sites observed in same read pair

Are Hi-C graphs are better connected than mate-pair??

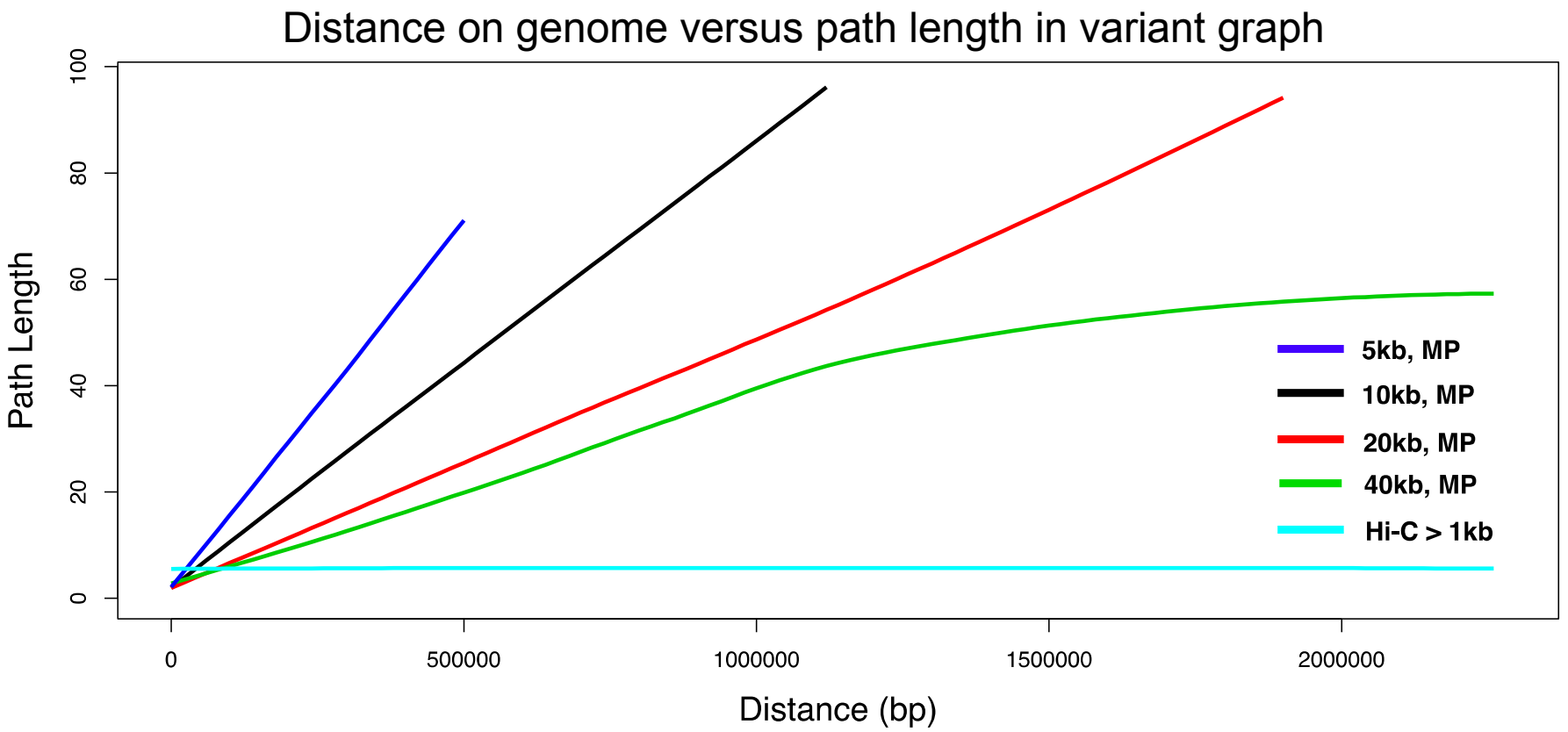
	5k MP	10k MP	20k MP	Hi-C
Nodes in largest connected component	6.2%	16.6%	32.4%	97.8%



Mate-pairs and paired-end reads make only “local” connections
Hi-C makes local and global connections

Beitel *et al* 2014 *PeerJ*.

Resolving strain differences: Hi-C versus mate-pair



Mate pair graph distances scale linearly.

Hi-C graphs are scale invariant.

Estimation error in probability of variant linkage grows with path length!

Beitel *et al* 2014 *PeerJ*.