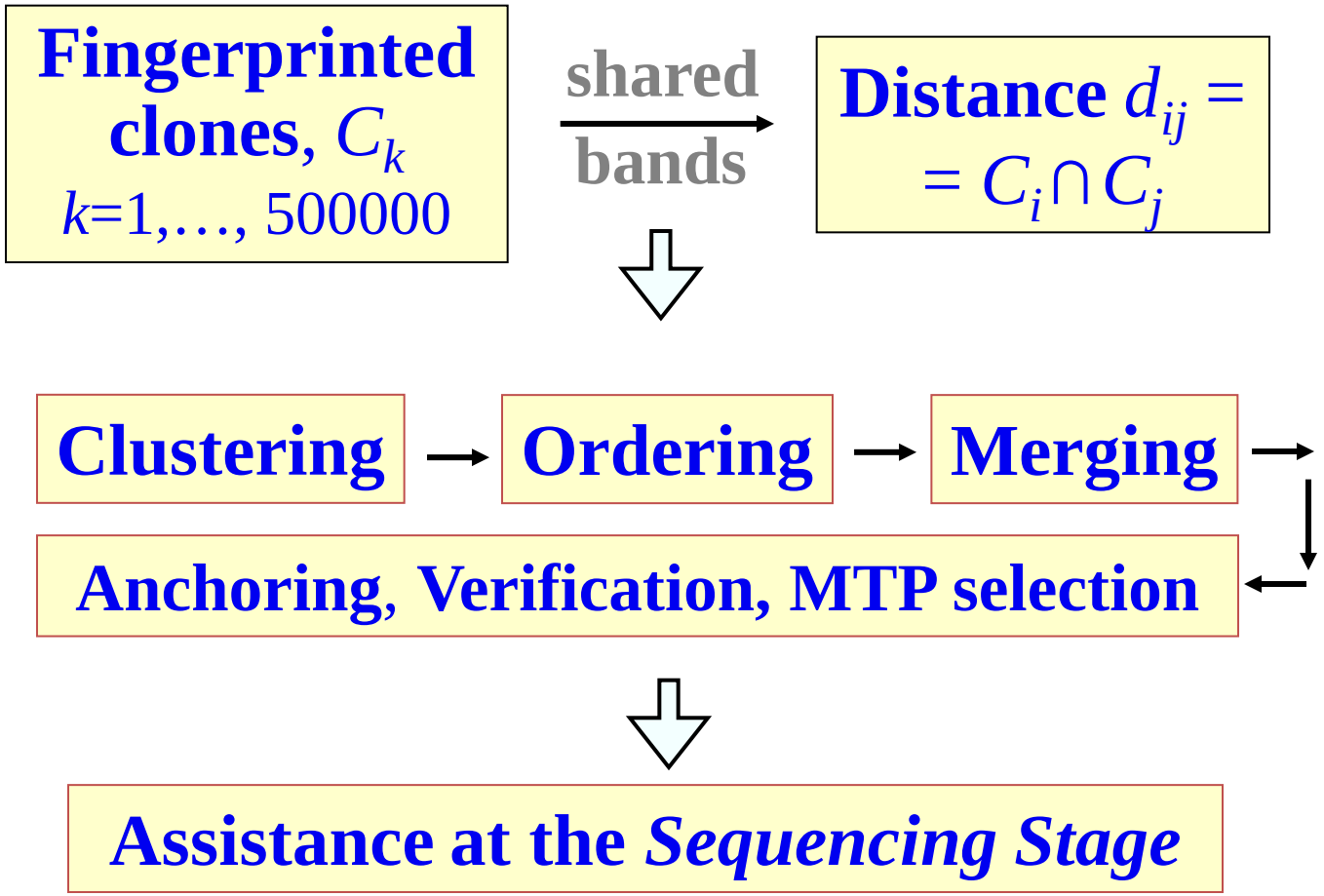
Using LTC Software for Wheat Physical Mapping: Increasing Contig Lengths and MTP Quality

A. Korol, Institute of Evolution, University of Haifa
korol@research.haifa.ac.il

See also poster P1130

PAGXXII, San Diego, January 2014

The major steps of physical mapping



Main difficulties in physical mapping

1. Chimerical clones
2. Low quality fingerprints
3. False clone overlaps due to repeats/duplications
4. 1-3 ➔ chimerical contigs
5. 1-4 ➔ problems in ordering
6. 1-5 ➔ problems in merging and anchoring
7. 3 & 5 ➔ gaps in MTP

⇨ **LTC**

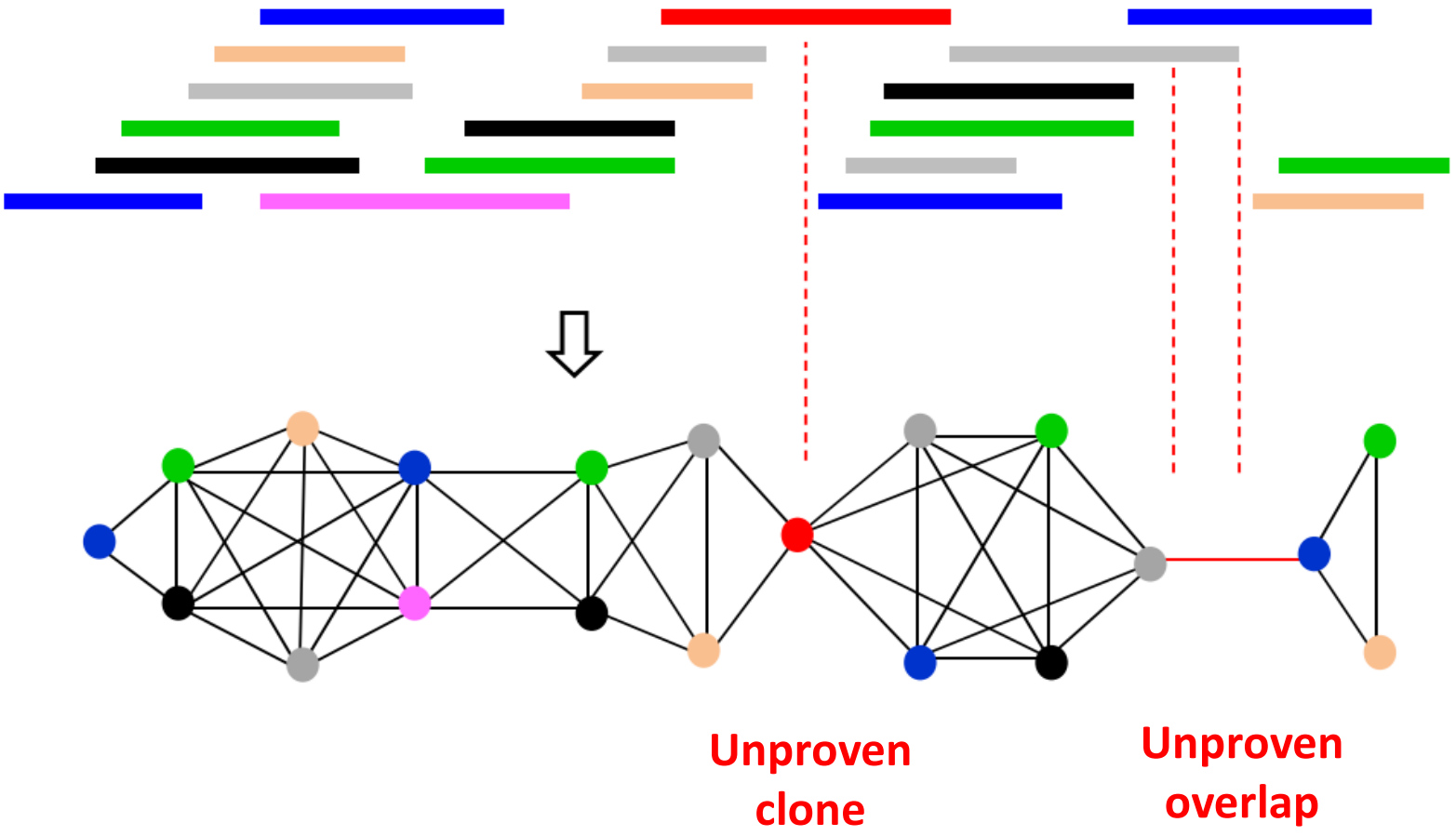
Contig assembly: LTC vs. FPC

- **Parallel clone overlaps** instead of consensus band/tag maps → more powerful detection of problematic clones and clone overlaps
- **Linear structure** of the net of significant clone overlaps → No contradictions of the contig topology with chromosome linear structure

➔ Longer and more reliable contigs

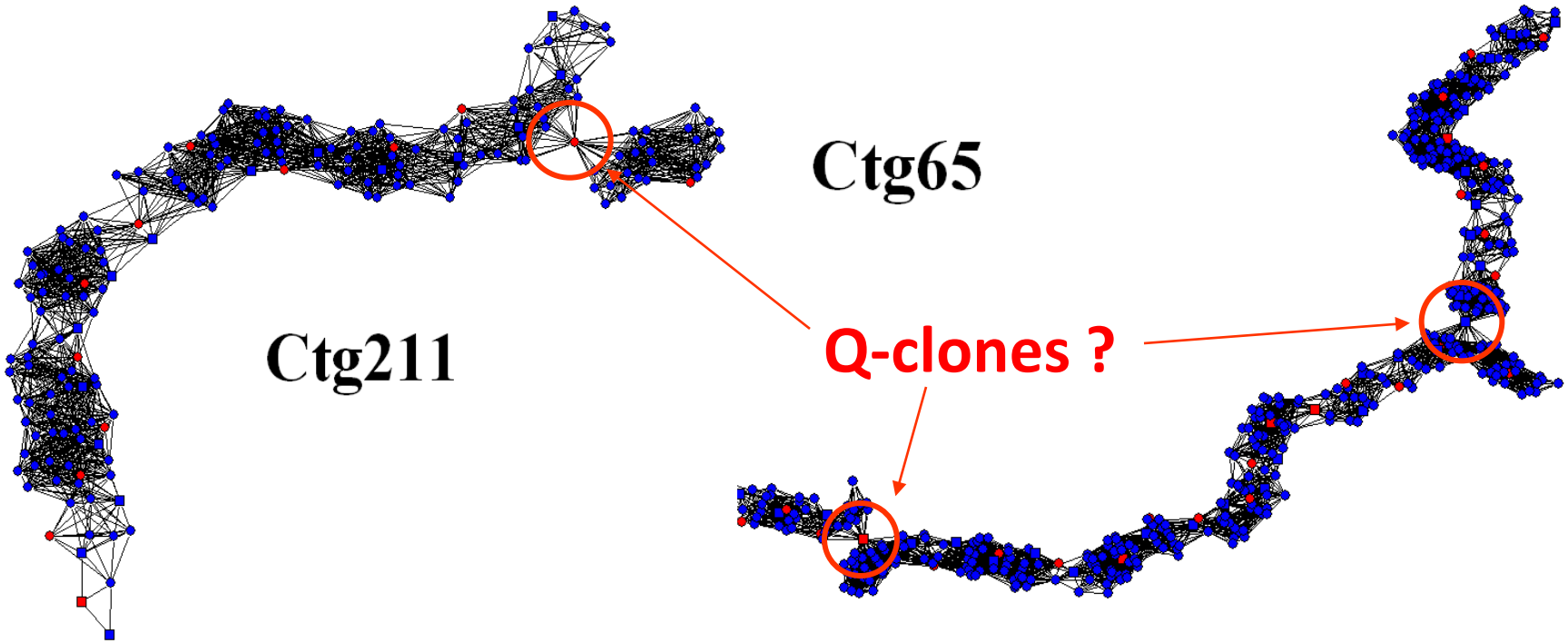
➔ Simpler anchoring

Net representation of clone overlaps



Testing FPC contig quality by using LTC

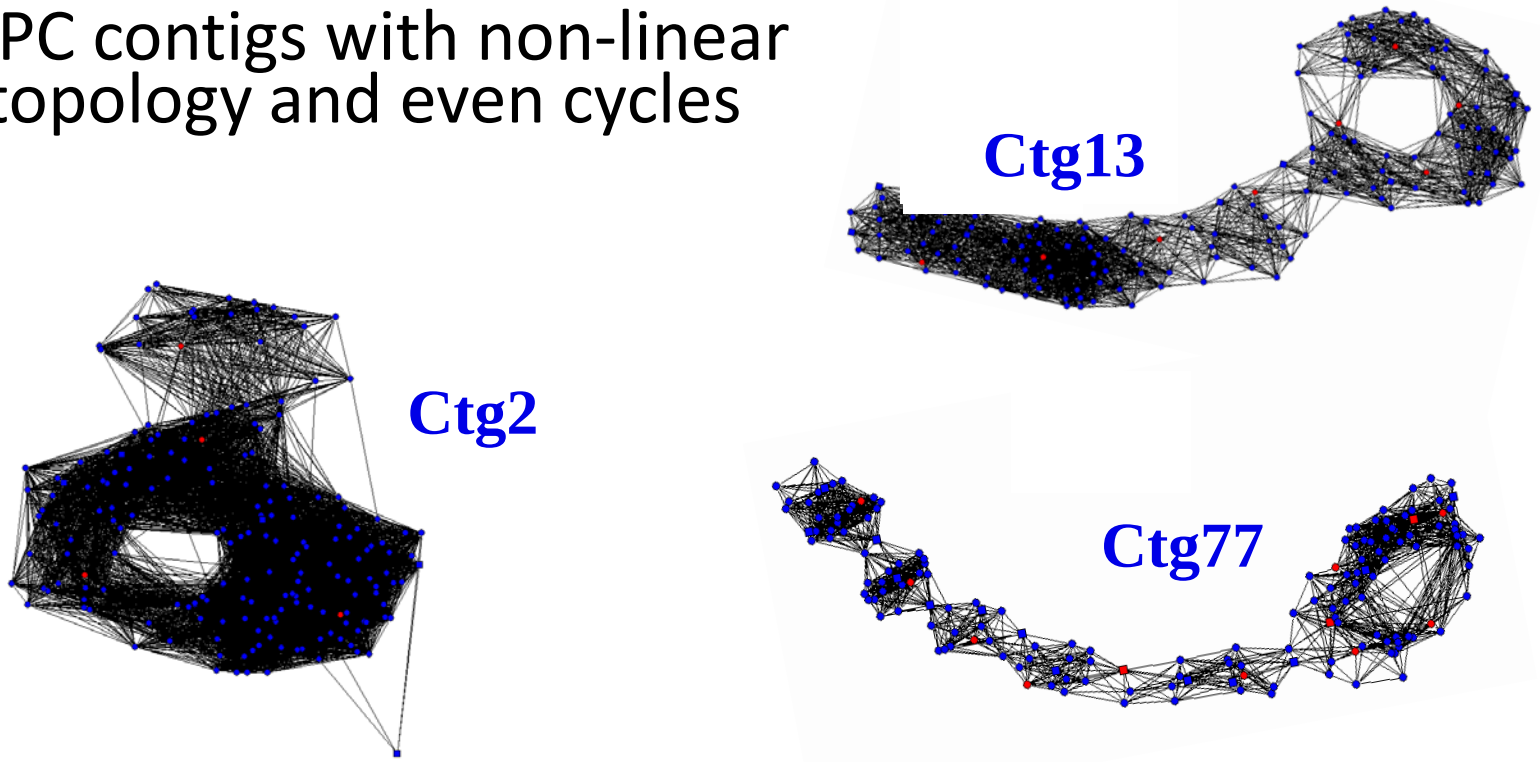
Some FPC contigs have non-linear topological structure inconsistent with chromosome linear structure:



Vertices represent the clones; **edges** represent the significant overlaps (with cutoff 1e-25 Sulston score)

Testing FPC contig quality by using LTC

FPC contigs with non-linear topology and even cycles



Edges represent significant overlaps (with cutoff 1e-25 Sulston score). Increasing the stringency up to e-75

does not help in non-trivial linearization!

Scaffolding of physical contigs

- Visual and analytical control of the net of significant clone overlaps
- Coordinating of scaffolding with anchoring

➔ Long well anchored physical scaffolds

Example: wheat 1BS (314 Mb, HICF, x15, ~50,000 BACs)

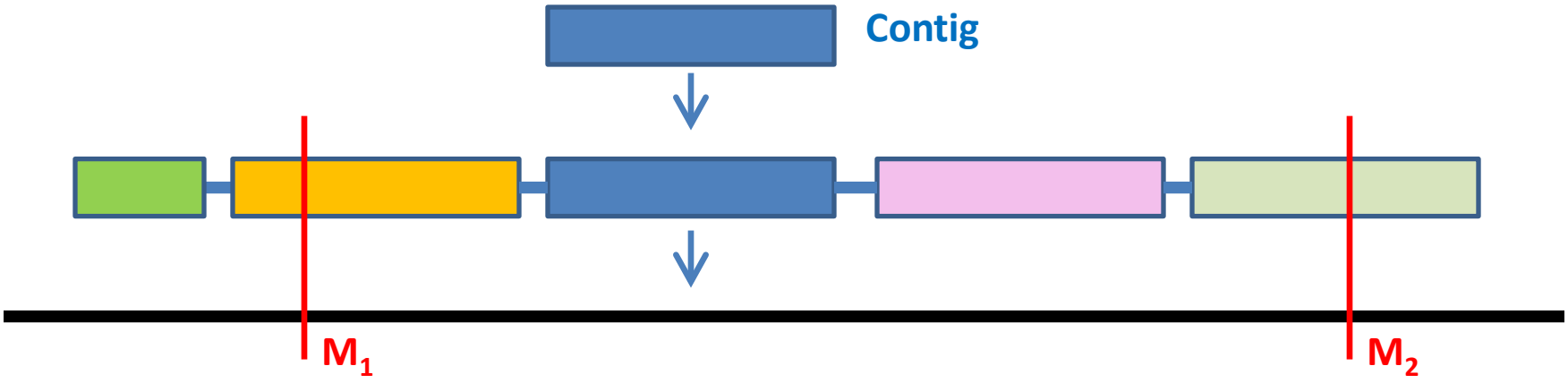
	FPC	LTC contigs	LTC scaffolds
Clones in contigs (≥6)	34,104	33,846	34,027
Longest contig (Mb)	4.7	**7.0**	**20.9**
N50 (Mb)	1.0	**2.4**	**8.5**
L50 (contigs)	81	**35**	**11**

Raats et al. *Genome Biology 2013,* ***14:R138***

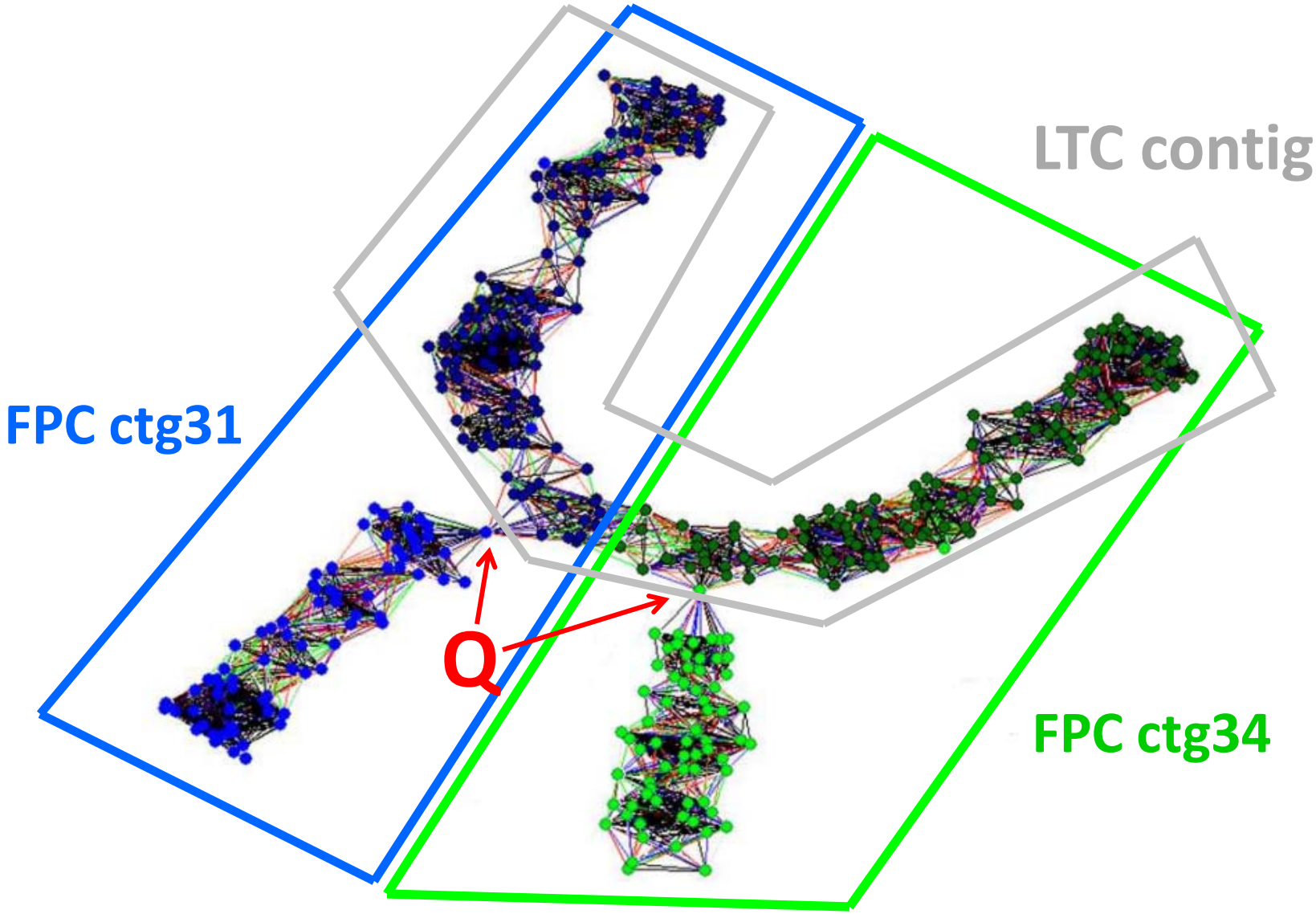
Anchoring of long contigs

- Much less markers are needed
- Especially useful for regions with suppressed recombination, e.g., “near” the centromeres
- More effective contig orientation in chromosomes

Scaffolds → possible anchoring and orientation even for contigs having no markers



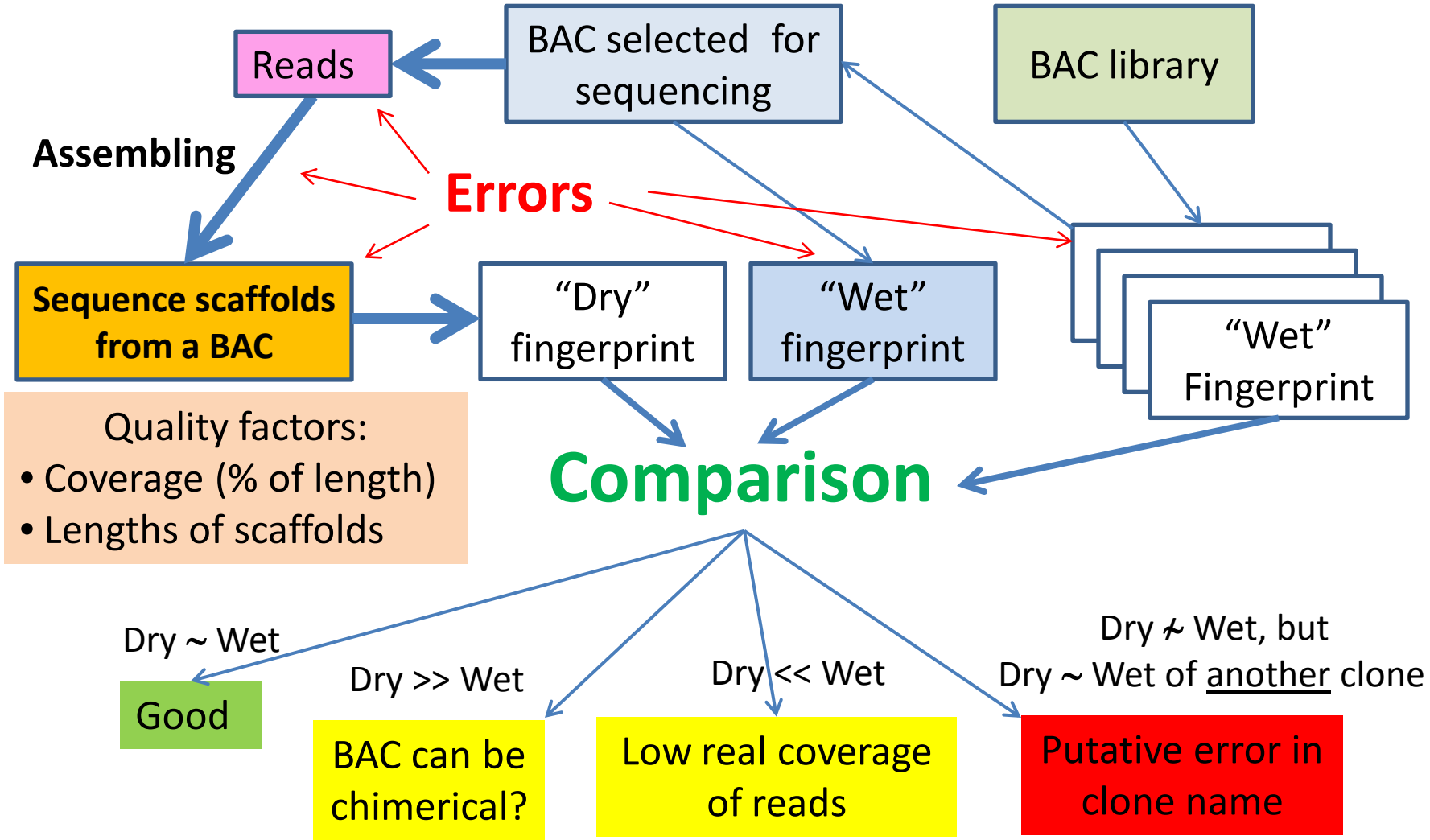
LTC scaffolds vs. FPC contigs



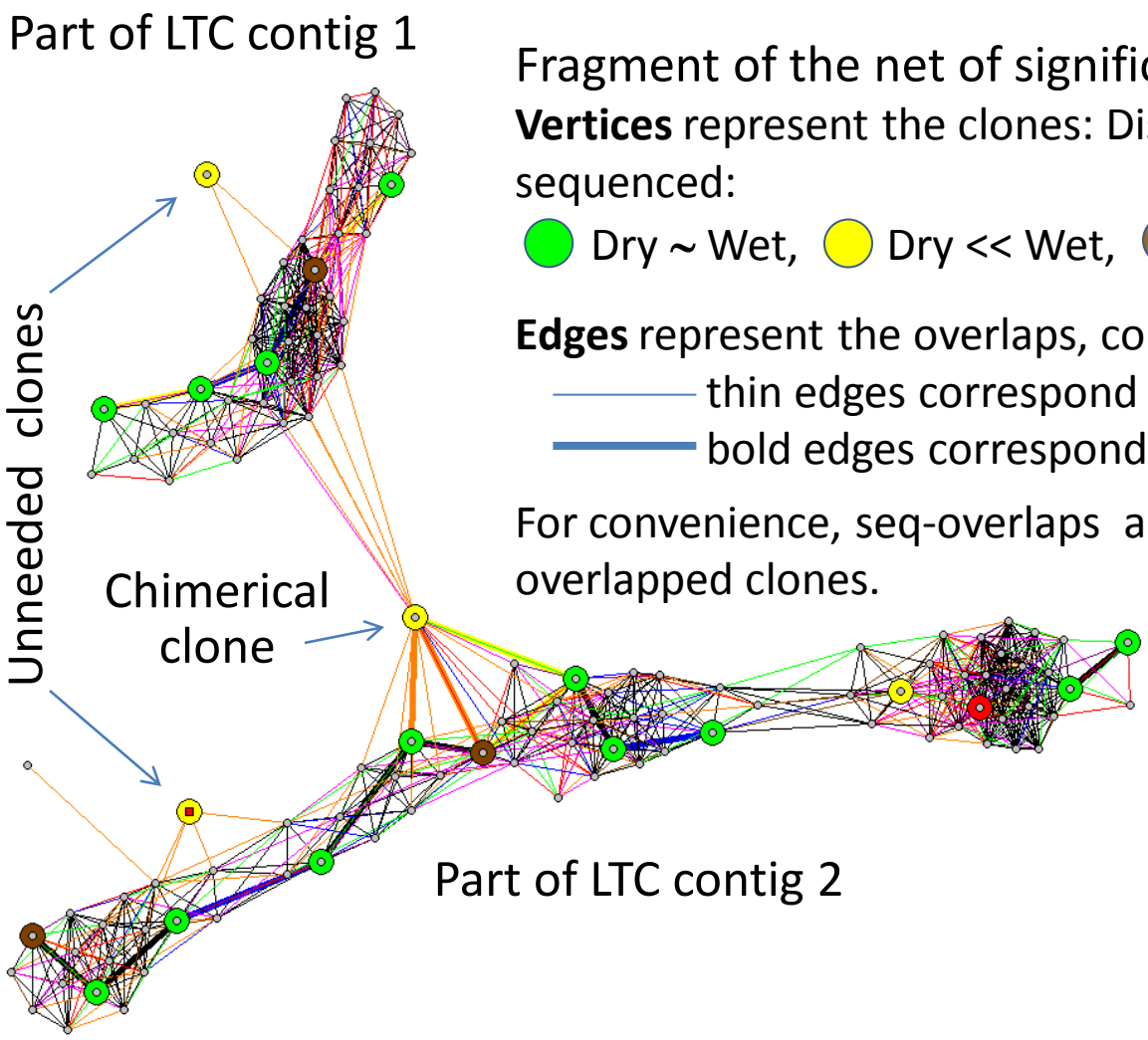
Selecting clones for sequencing by LTC

- Possibility to give priority to previously selected MTP clones (for anchoring or for BAC-end sequencing)
- Larger (more sure) overlaps of neighbor clones to avoid non-significant overlaps at sequence level in highly repeated genomes → **less gaps**
- Reducing the risk of errors caused by Q-clones and false clone overlaps → **more reliable MTP**
- Supplementing the list of MTP clones by potential "bridges" for end-to-end merging → **longer contigs**

Controlling the sequencing quality



LTC control of MTP clone-overlaps at sequence level



Fragment of the net of significant clone overlaps (7BS data)

Vertices represent the clones: Disk indicate that the clone was sequenced:

Dry ~ Wet, Dry << Wet, Dry >> Wet, Dry ≁ Wet

Edges represent the overlaps, color reflects significance:

thin edges correspond to HICF-based overlaps

bold edges correspond to seq-based overlaps

For convenience, seq-overlaps are shown only for HICF-overlapped clones.

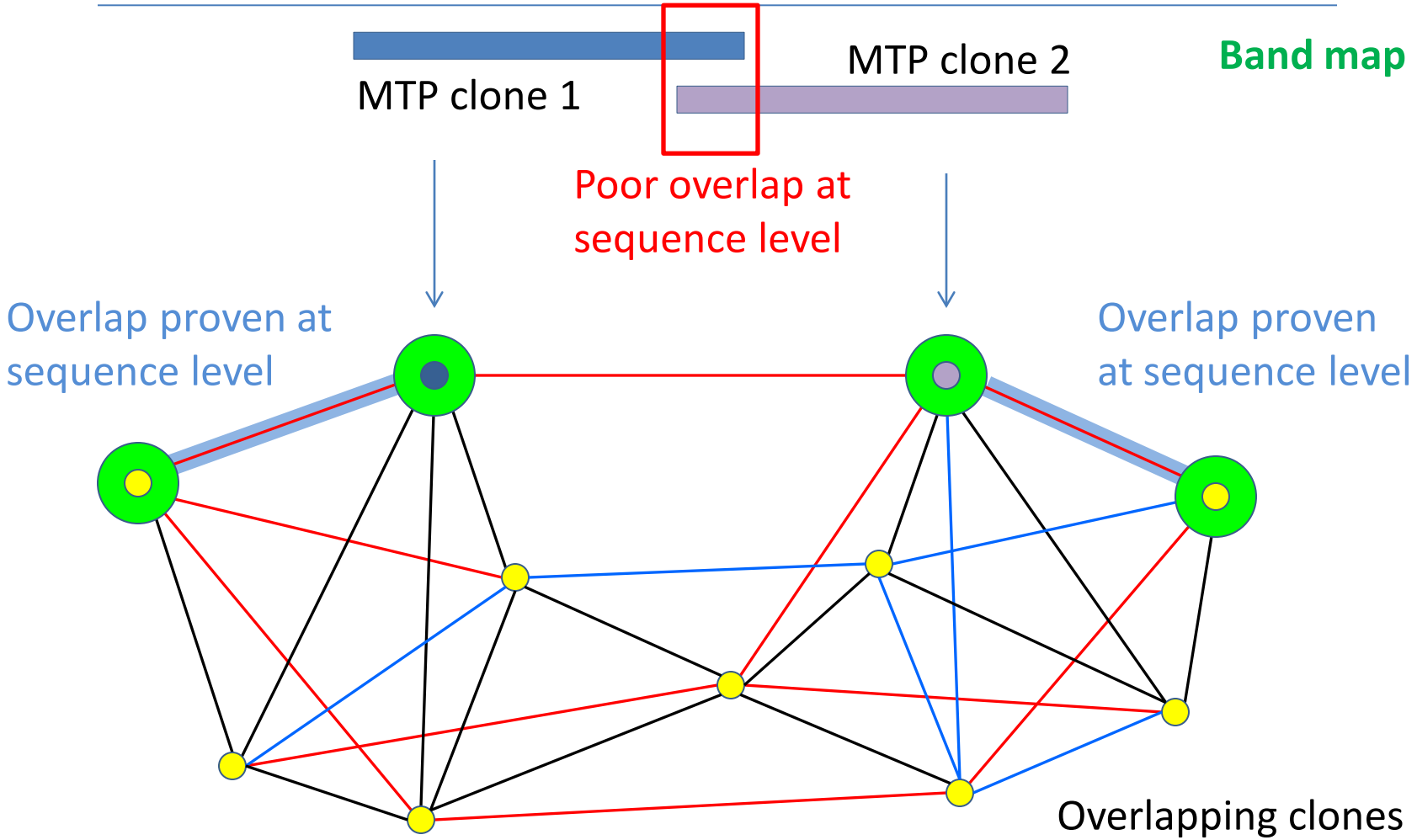
LTC candidate solutions to cure the detected **sequence** gaps

- **Check the physical contig:** a gap can be a result of error(s) in physical contig assembly
- **Check overlaps** in fingerprints
- **Check sequence quality**: coverage, length and correspondence of wet and dry fingerprints
- **Add clones** to connect the sides of the gap via significant fingerprint-based overlaps
- If well sequenced clones appeared to overlap on fingerprint but not sequence level, try to **increase cutoff at the fingerprint level**

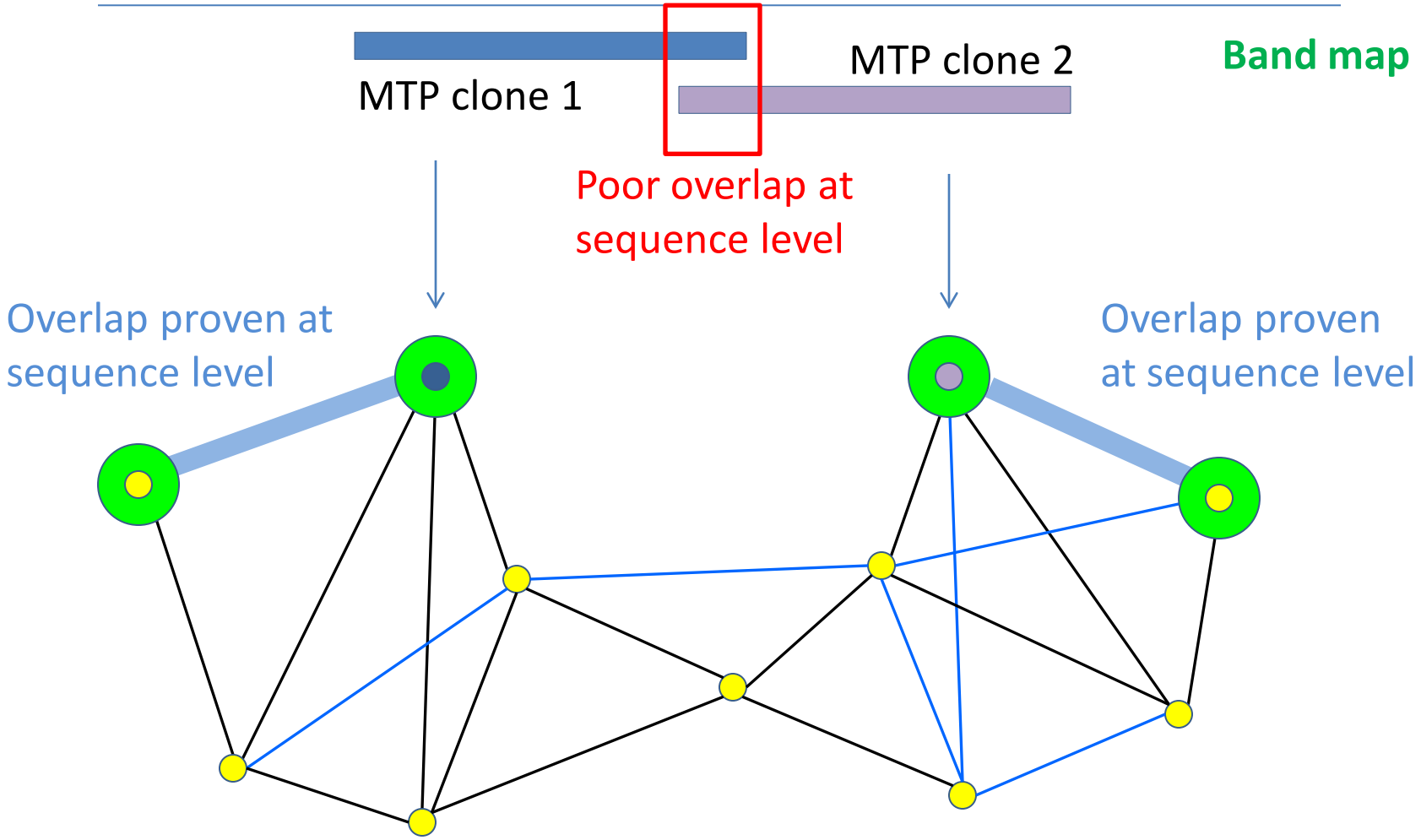
LTC candidate solutions to cure the detected **sequence** gaps

- **Check the physical contig:** a gap can be a result of error(s) in physical contig assembly
- **Check overlaps** in fingerprints
- **Check sequence quality**: coverage, length and correspondence of wet and dry fingerprints
- **Add clones** to connect the sides of the gap via significant fingerprint-based overlaps
- If well sequenced clones appeared to overlap on fingerprint but not sequence level, try to **increase cutoff at the fingerprint level**

Example of gap repairing

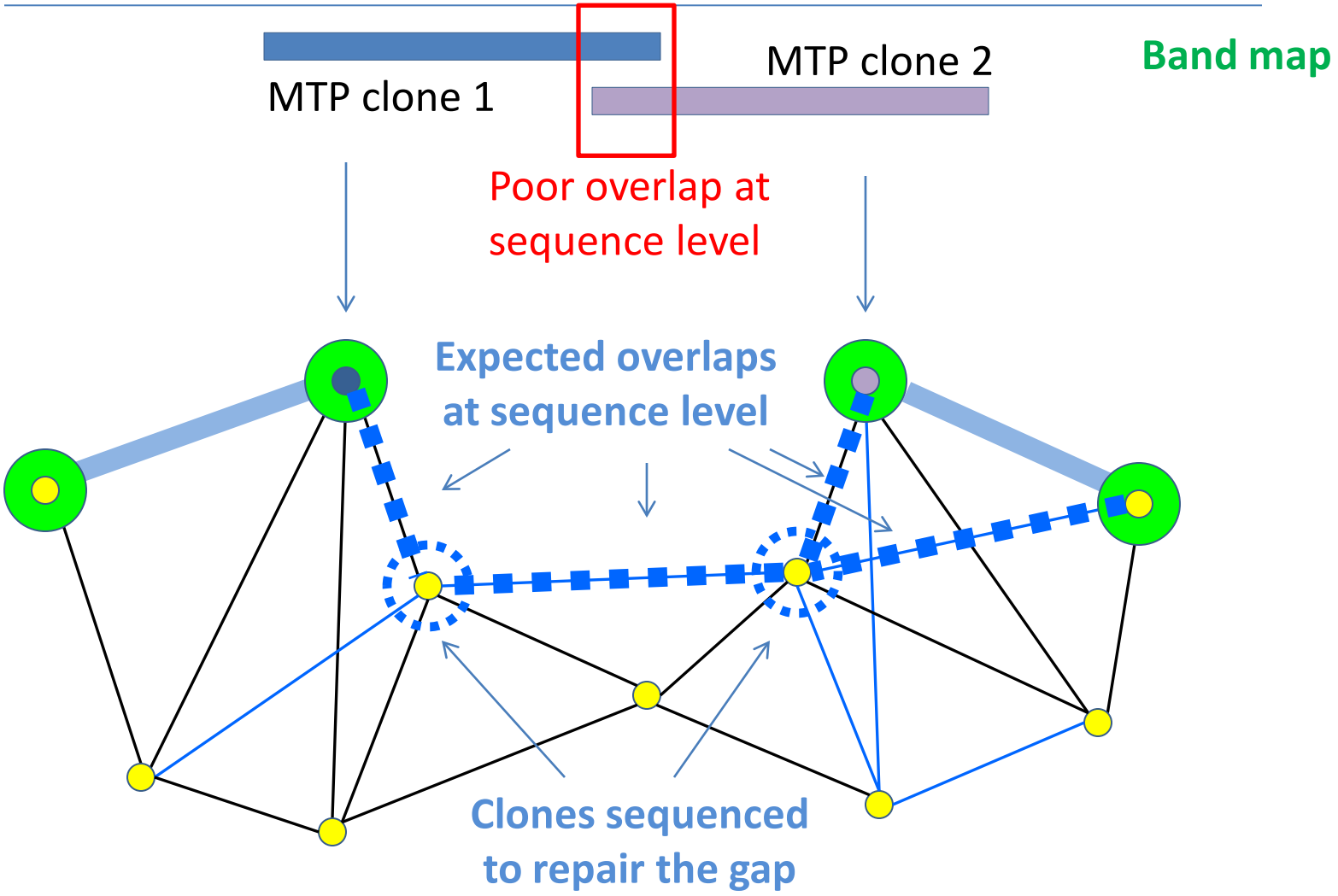


Example of gap repairing



Clone overlaps detected at increased cutoff stringency

Example of gap repairing



Some prospects

- Simplification of scaffolding of physical contigs coordinated with anchoring
- Optimization of MTP selection by taking into account clone length, clone overlaps and putative (calculated) local coverage and repetitiveness
- Orientation, ordering and merging of sequence scaffolds assisted by fingerprinting information from overlapped fingerprinted clones (even not yet sequenced)

Some prospects

- Simplification of scaffolding of physical contigs coordinated with anchoring
- Optimization of MTP selection by taking into account clone length, clone overlaps and putative (calculated) local coverage and repetitiveness
- Orientation, ordering and merging of sequence scaffolds assisted by fingerprinting information from overlapped fingerprinted clones (even not yet sequenced)

Example: sequencing of YrH52 and Yr15 region (1BS)

Candidate region:

Length ~6Mb

Covered by 104 overlapping MTP clones

Pooling of neighbor MTP clones :

23 pools instead 104

→ lower cost of sequencing stage

Sequencing by MiSeq (x450 coverage)

Orientation, ordering and merging of sequence scaffolds

Sequence contig assembly (using EDENA):

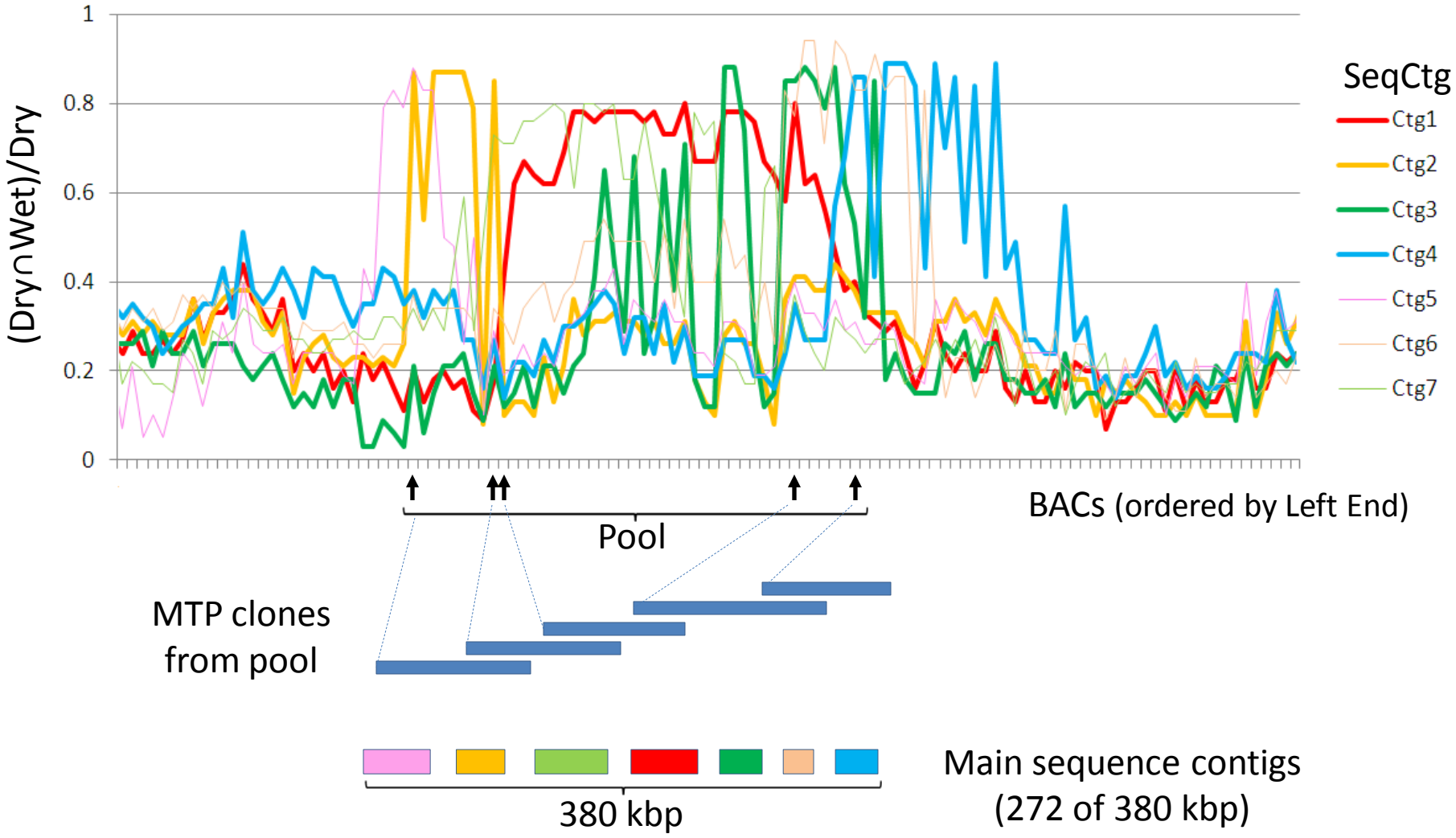
- 9-56 sequence contigs per pool
- Average total length of contigs per pool ~ 333 kbp
- **Only few "main" contigs** (longer than 15 kbp)

Sequence contigs → *in silico* fingerprinting

↓

Comparison with clones from physical scaffold (not MTP only)

Ordering and orientation of sequence contigs within pool



Acknowledgements

- Etienne Paux
- Catherine Feuillet

- Kellye Eversole

- Simen Sandve
- Bujie Zhan
- Tatiana Belova
- Odd-Arne Olsen

Norwegian University of Life Sciences

- Zeev Frenkel
- Vladimir Glikson

- Dina Raats
- Tzion Fahima