

xia2 & pipelines

CCP4 developers meeting 2012

Graeme Winter

Diamond Light Source

21 March 2012



Overview

- xia2 developments
- Removal of frame count limitations
- Current state of multi-crystal analysis
- Plans for multi-crystal analysis
- Fast EP



xia2 developments

- xia2 for multi-axis goniometers and small unit cells
- Bugs / fixes
- Using Aimless
- P1 integration
- ~~Scaling XDS data with Scala / Aimless~~



Removal of frame count limitations

- Use Aimless in place of Scala
- Use Pointless in place of Reindex¹
- Run -3daii or -3da
- No more batch limitations

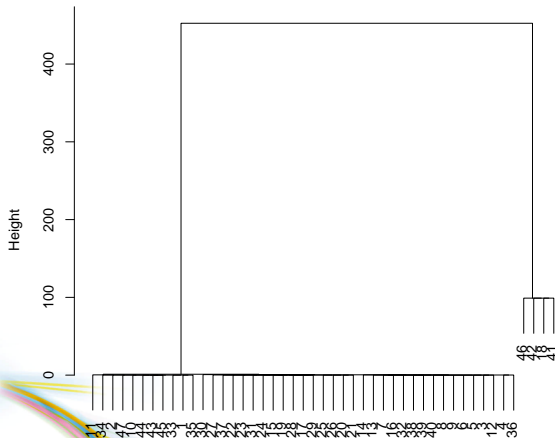
¹This is still not as efficient as it could be - but it works

Plate data

- Data from Diamond I04-1
- 110 × 100 image sweeps (20 degree)
- Processed with xia2 -3daii -failover -spacegroup ... -cell ...
- ... crashed (this is to be expected!)
- In XSCALE output analyse $CC(i, j)$, scale to $d = \frac{1}{CC} - 1$ - useful distances
- Get dendrogram

Results

Cluster Dendrogram



d
hclust (*, "ward")

Next steps

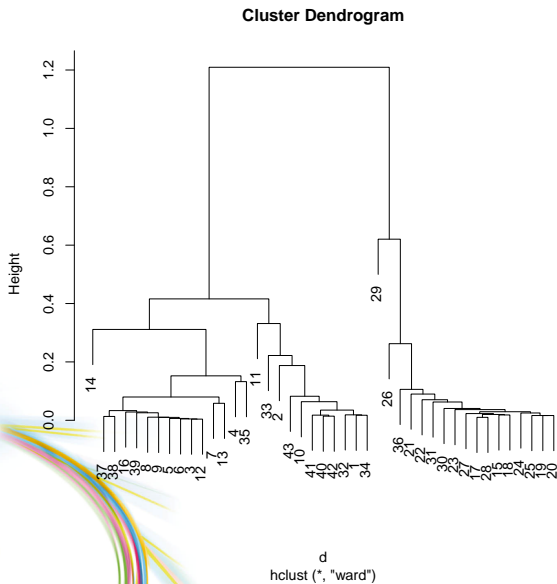
- Remove useless sweeps, run again
- This process is partially automated
- `xia2 -3daii (etc.) -xinfo mod.xinfo`
- Runs now to completion



For AUTOMATIC/DEFAULT/NATIVE

High resolution limit	1.77	7.93	1.77
Low resolution limit	39.42	39.42	1.82
Completeness	88.7	99.5	19.7
Multiplicity	39.2	46.8	1.8
I/sigma	21.0	35.1	3.1
Rmerge	0.275	0.261	0.204
Rmeas(I)	0.279	0.264	0.281
Rmeas(I+/-)	0.280	0.265	0.268
Rpim(I)	0.037	0.039	0.168
Rpim(I+/-)	0.049	0.047	0.172
Wilson B factor	22.655		
Total observations	429848	8325	318
Total unique	10962	178	174

Results



Cluster 1: 37, 38, 16, 39, 8, 9, 5, 6, 3, 12

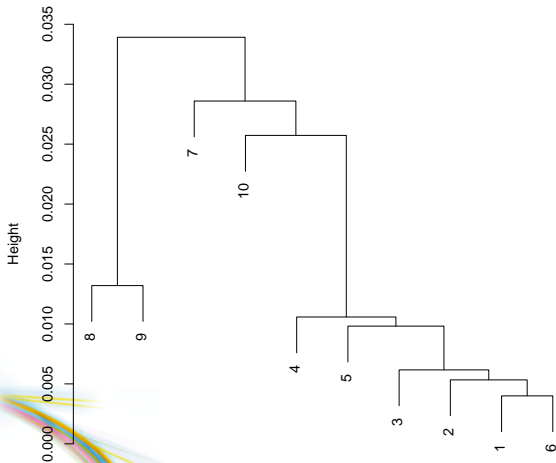
- Use Re-Xinfo to take selected runs from dendrogram, x1335 output and automatic.xinfo to generate new xinfo
- ... and run *again*



High resolution limit	1.74	7.79	1.74
Low resolution limit	39.42	39.42	1.79
Completeness	84.2	93.4	16.3
Multiplicity	11.2	12.3	1.5
I/sigma	17.4	31.9	2.5
Rmerge	0.099	0.068	0.213
Rmeas(I)	0.105	0.072	0.334
Rmeas(I+/-)	0.105	0.072	0.292
Rpim(I)	0.026	0.019	0.211
Rpim(I+/-)	0.034	0.022	0.199
Wilson B factor	22.448		
Total observations	121810	2097	237
Total unique	10858	170	153

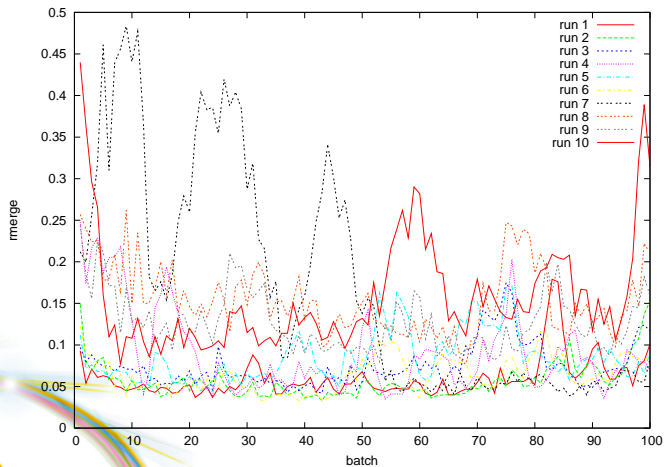
Results

Cluster Dendrogram



d
hclust (*, "ward")

Results



Next steps: April 2012 onwards

- Automate this procedure, allow user to specify requirements
- Include BLEND analysis: complementary information
- Code up own CC calculation: can run from Aimless output
- Automate R_{merge} vs. BATCH vs. run analysis
- Work on simply processing more tricky / weaker data



Fast EP

- Experimental phasing in the spirit of Fast DP
- Uses shelxc / d_mp / e
- Extensive use of CCTBX for analysis
- Can give maps within 2 minutes of experiment in favourable cases - DP 1 minute, EP 1 minute
- Takes no information, searches everything (spacegroups, sites, hands)

Brute force / power

- Assess with CCTBX code
- Use shelxc to get ins file, hkl file
- Modify using CCTBX
- Run shelxd_mp on cluster: up to 480 x 3GHz cpu's
- Sort results on CFOM, find best, get number sites
- Run 22 shelxe runs: 2 hands, 11 solvent fractions, score on pseudo-free CC

Testing / proof of principle

- Data from Raj / Pavel
- Trivial, tricky, impossible examples
- Immediate response: phasing quality in shelxe depends on resolution
- Substructure solution much less limited



Limitations

- Shelxe limited for phasing: use something more potent?
- Hand determination



Value

- Fast feedback for phasing experiments
- Solve structures automatically for users
- Has already cracked a useful novel structure

