



**CBRS *Chlamydiae*
community re-annotation**



Session schedule

- Intro and automatic annotation (T. Weinmaier)
 - Discussion
- Nomenclature
 - Discussion
- Manual refinement
 - Discussion
- Submission / Publication
 - Discussion



Session schedule

- **Intro and automatic annotation (T. Weinmaier)**
 - Discussion
- Nomenclature
 - Discussion
- Manual refinement
 - Discussion
- Submission / Publication
 - Discussion

Concept for *Chlamydiae* community re-annotation

Thomas Weinmaier

Chlamydial Basic Research Society meeting,
San Antonio, Texas

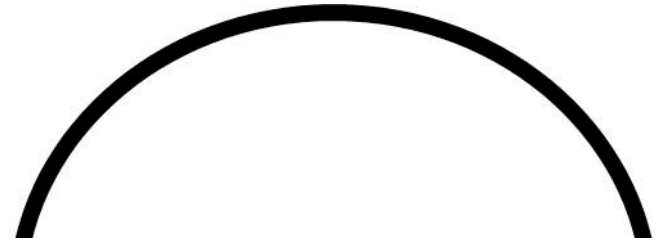
Prototypic workflow for genome projects

- Sequencing
- Assembly
- Primary annotation
- Functional annotation



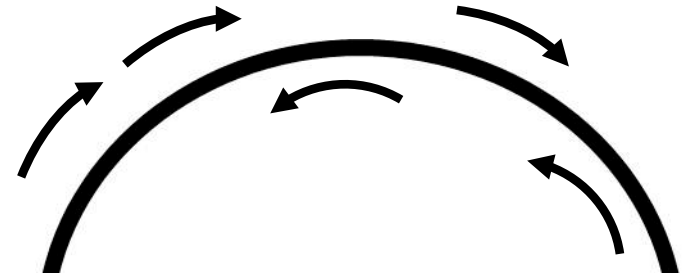
Prototypic workflow for genome projects

- Sequencing
- **Assembly**
- Primary annotation
- Functional annotation



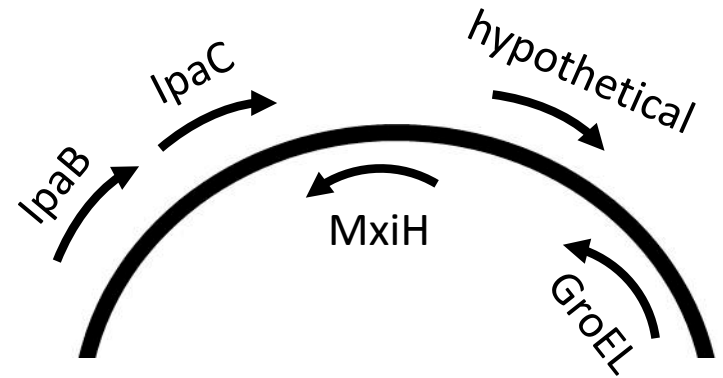
Prototypic workflow for genome projects

- Sequencing
- Assembly
- **Primary annotation**
- Functional annotation



Prototypic workflow for genome projects

- Sequencing
- Assembly
- Primary annotation
- **Functional annotation**

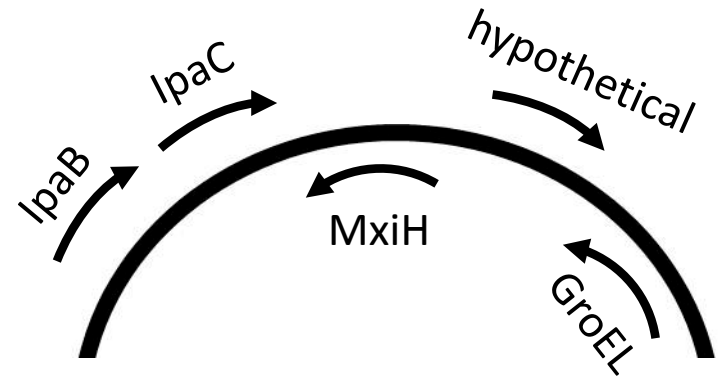


Prototypic workflow for genome projects

- Sequencing
- Assembly
- Primary annotation
- Functional annotation

Annotation drawbacks:

- No standard procedure
- Later not updated



Comparative genomics

- *C. pneumoniae* clinical isolates: AR39, CWL029, J138 and TW183
- Genome sizes ~1,23 Mb
- ≤ 6000 nucleotides different (99.5% identical)

Annotated genes in Genbank

Isolate	# Genes
AR39	1110
CWL029	1052
J138	1069
TW183	1113

Genes without ortholog

Isolate	AR39
AR39	-
CWL029	76
J138	77
TW183	72

BLAST search at NCBI

Query: chlamydial protease like activity factor (CPAF)

[*Waddlia chondrophila* WSU 86-1044]

Search: BLASTP against NCBI RefSeq database

BLAST search at NCBI

Query: chlamydial protease like activity factor (CPAF)
[Waddlia chondrophila WSU 86-1044]

Search: BLASTP against NCBI RefSeq database



Description

[chlamydial protease-like activity factor \(CPAF\) \[Waddlia chondrophila WSU 86-1044\]](#)

[putative chlamydial protease-like activity factor \[Parachlamydia acanthamoebae str. Hall's coccus\] >ref](#)

[protease-like activity factor \[Protochlamydia amoebophila UWE25\]](#)

[hypothetical protein CAB712 \[Chlamydophila abortus S26/3\]](#)

[hypothetical protein CAB1_0732 \[Chlamydophila abortus LLG\]](#)

BLAST search at NCBI

Query: chlamydial protease like activity factor (CPAF)
[Waddlia chondrophila WSU 86-1044]

Search: BLASTP against NCBI RefSeq database



Description

[chlamydial protease-like activity factor \(CPAF\) \[Waddlia chondrophila WSU 86-1044\]](#)

[putative chlamydial protease-like activity factor \[Parachlamydia acanthamoebae str. Hall's coccus\] >ref](#)

[protease-like activity factor \[Protochlamydia amoebophila UWE25\]](#)

[hypothetical protein CAB712 \[Chlamydophila abortus S26/3\]](#)

[hypothetical protein CAB1_0732 \[Chlamydophila abortus LLG\]](#)

83 chlamydial genomes

Problem:

- Different annotation strategies
- No update after submission

83 chlamydial genomes

Problem:

- Different annotation strategies
- No update after submission

Re-annotation goal:

- Consistency
- Currentness

83 chlamydial genomes

Problem:

- Different annotation strategies
- No update after submission

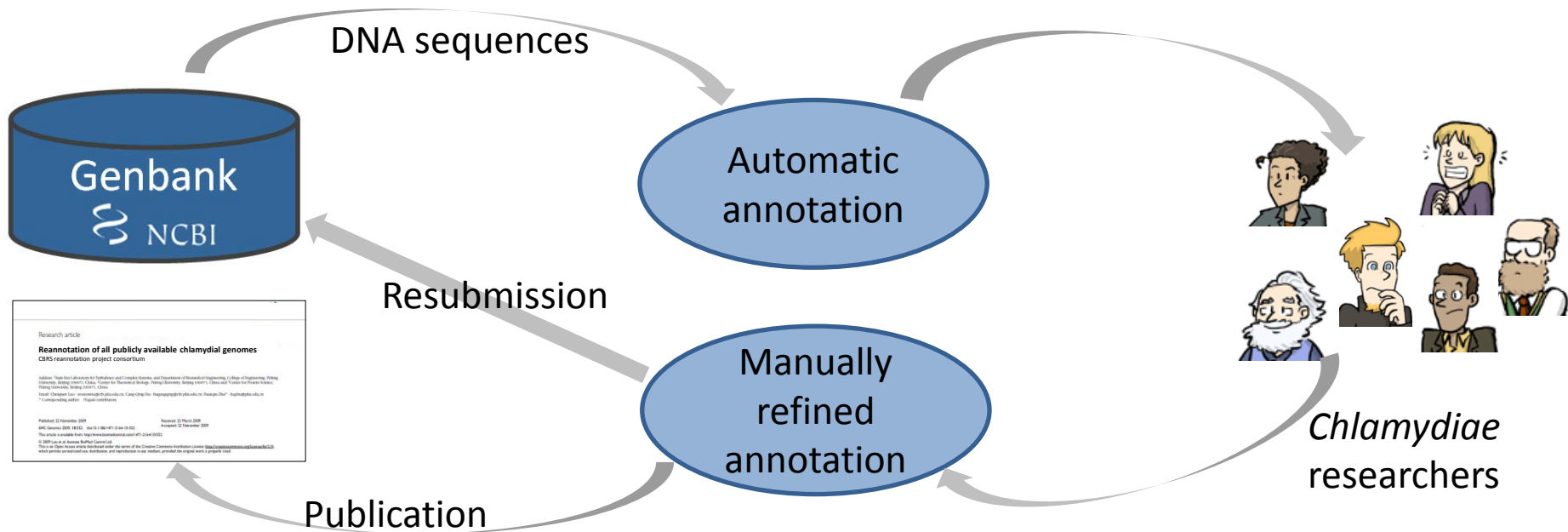
Re-annotation goal:

- Consistency
- Currentness

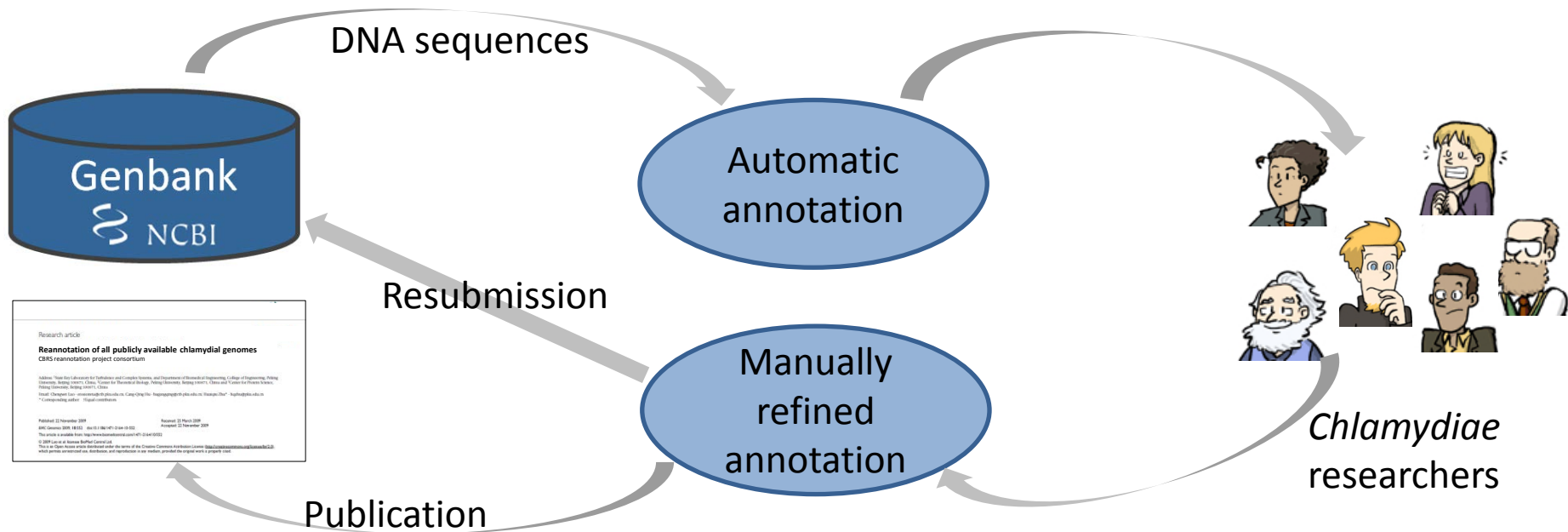
Solution:

- Consistent automatic re-annotation
- Manual refinement

Proposed re-annotation strategy

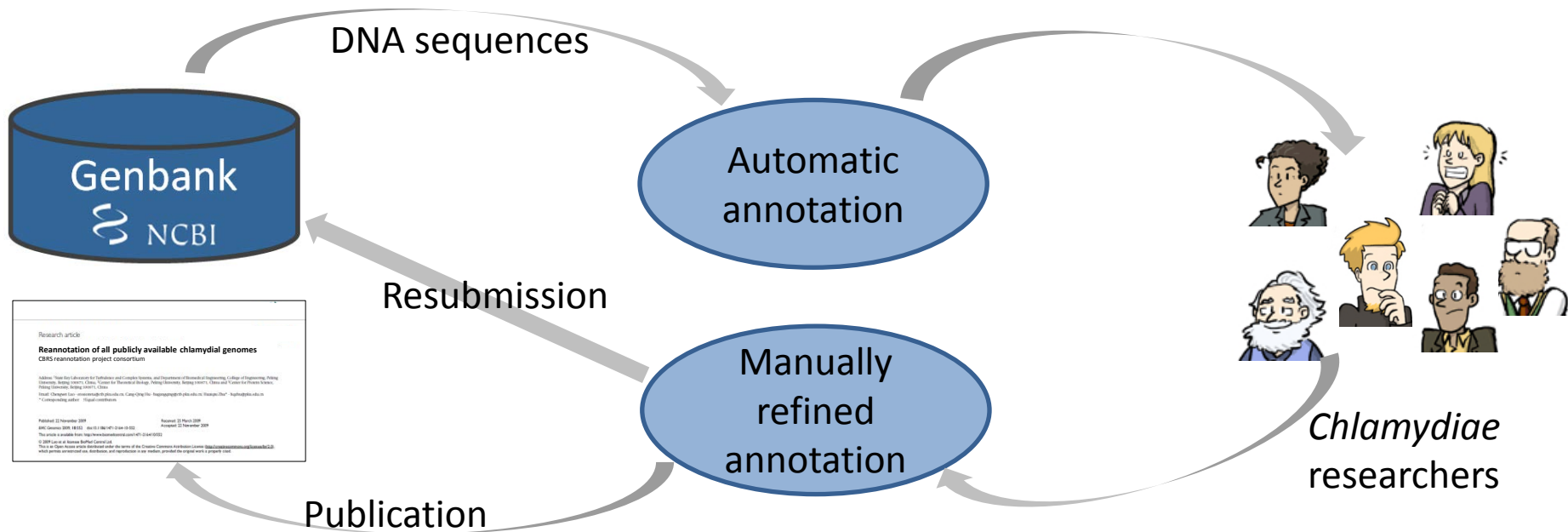


Proposed re-annotation strategy



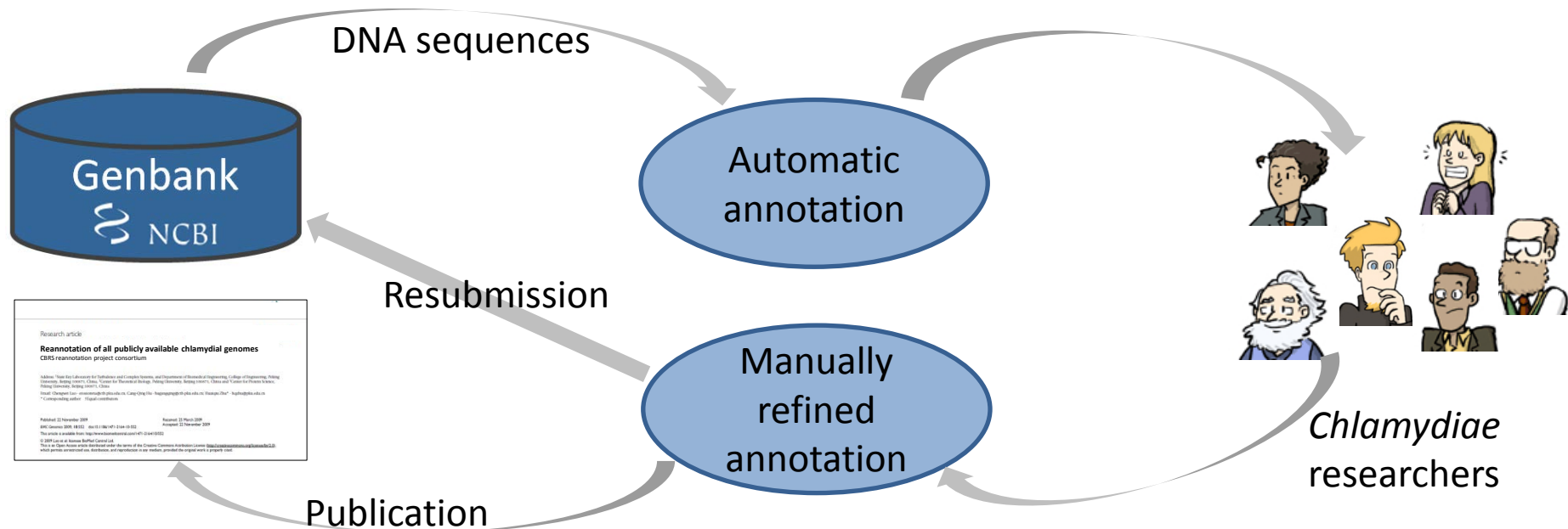
- DNA sequences from Genbank
- Automatic re-annotation

Proposed re-annotation strategy



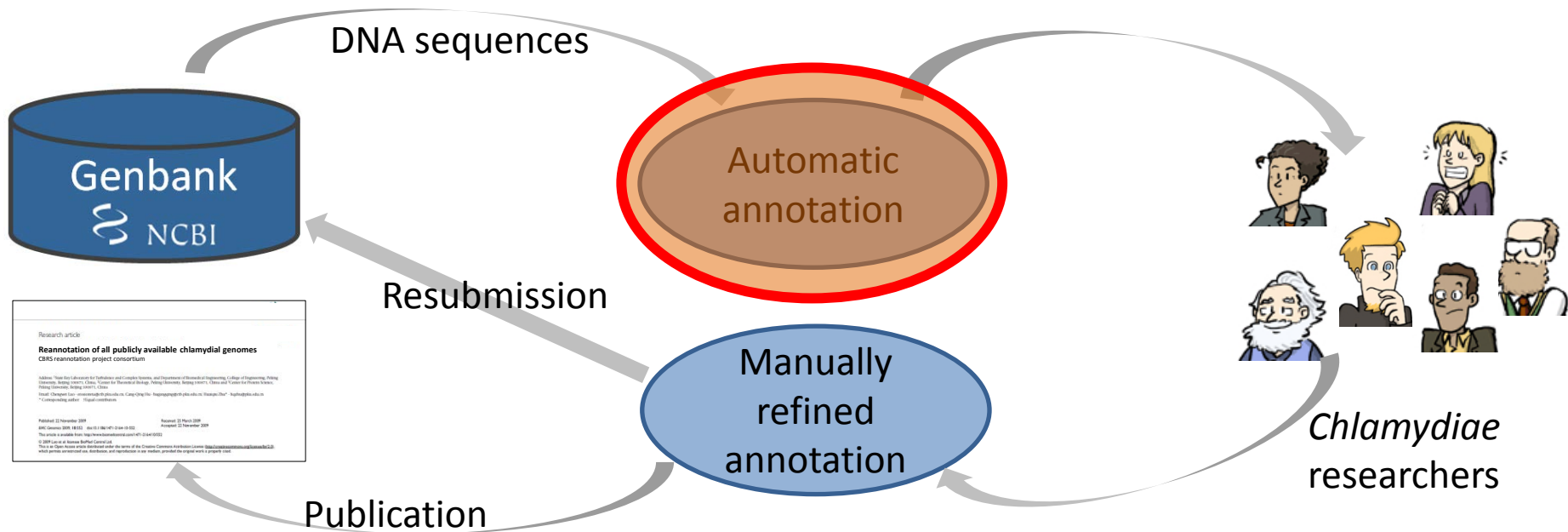
- DNA sequences from Genbank
- Automatic re-annotation
- Manual refinement

Proposed re-annotation strategy



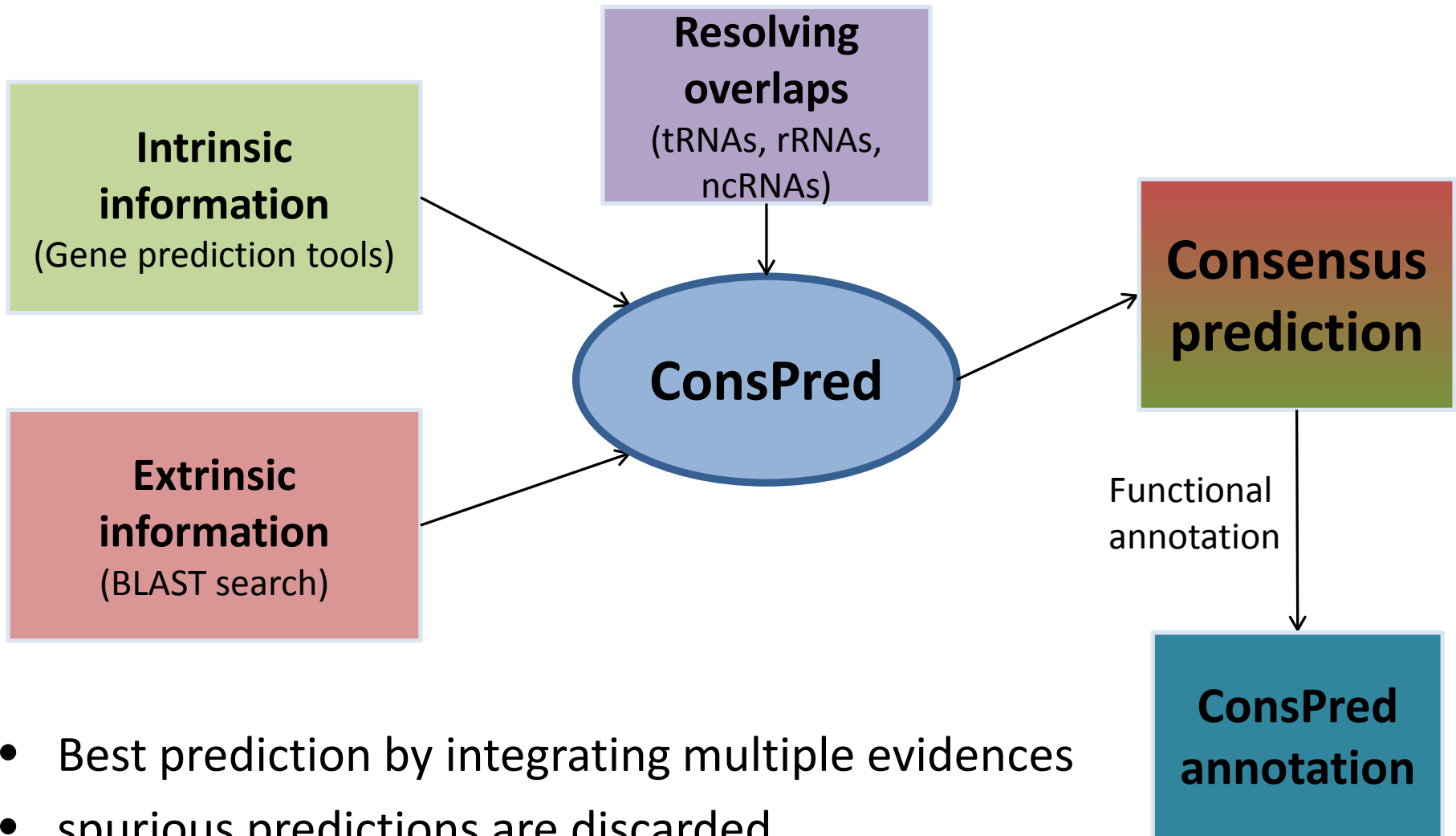
- DNA sequences from Genbank
- Automatic re-annotation
- Manual refinement
- Resubmission to Genbank
- Publication

Proposed re-annotation strategy



- DNA sequences from Genbank
- Automatic re-annotation
- Manual refinement
- Resubmission to Genbank
- Publication

Automatic annotation software



- Best prediction by integrating multiple evidences
- spurious predictions are discarded

Comparative genomics after re-annotation

- *C. pneumoniae* clinical isolates: AR39, CWL029, J138 and TW183
- Genome sizes ~1,23 Mb
- ≤ 6000 nucleotides different (99.5% identical)

Annotated genes from ConsPred

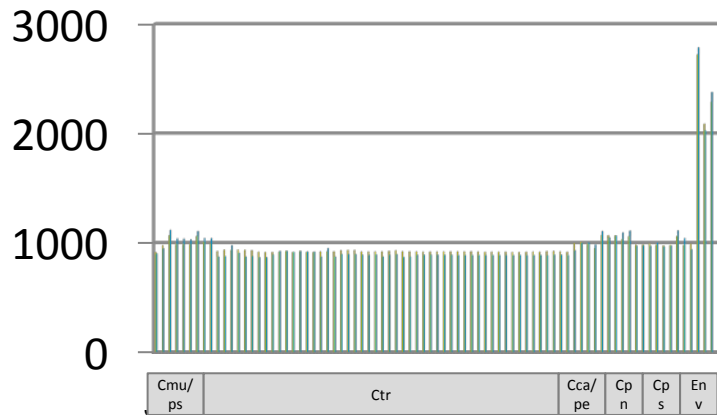
Isolate	# Genes
AR39	1110 1072
CWL029	1052 1072
J138	1069 1071
TW183	1113 1063

Genes without ortholog

Isolate	AR39
AR39	-
CWL029	76 16
J138	77 17
TW183	72 24

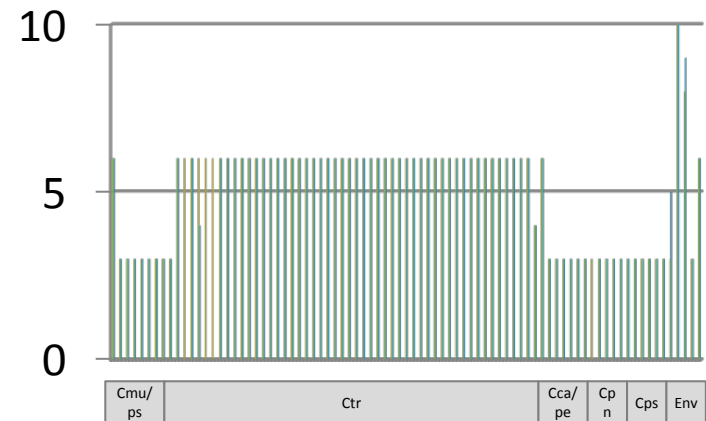
Gene counts

genes

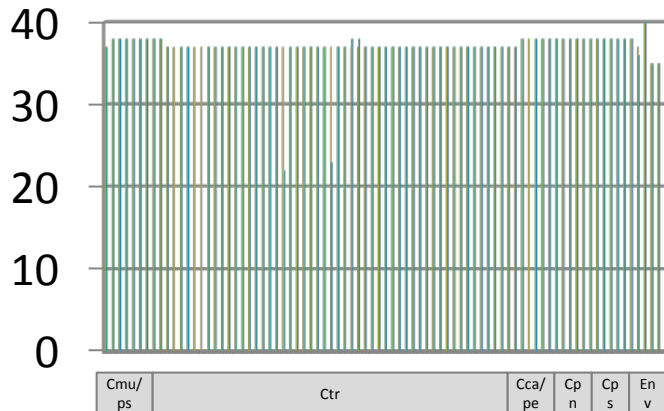


rRNAs

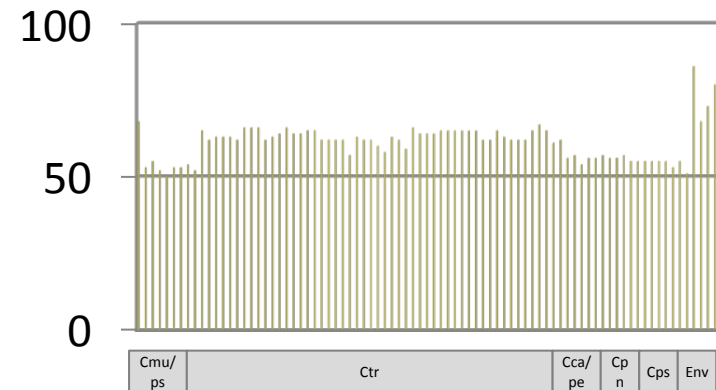
■ ConsPred
■ Genbank



tRNAs



ncRNAs





Session schedule

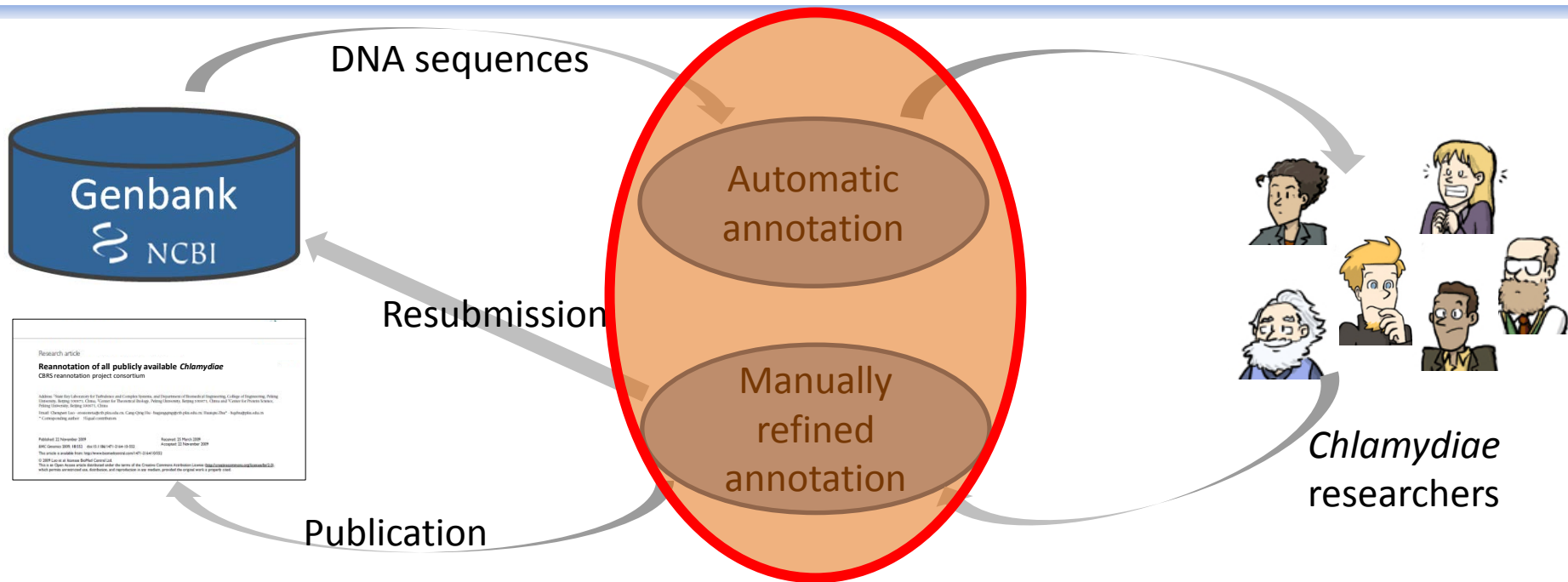
- Intro and automatic annotation (T. Weinmaier)
 - Discussion
- Nomenclature
 - Discussion
- Manual refinement
 - Discussion
- Submission / Publication
 - Discussion



Session schedule

- Intro and automatic annotation (T. Weinmaier)
 - Discussion
- **Nomenclature**
 - Discussion
- Manual refinement
 - Discussion
- Submission / Publication
 - Discussion

Proposed re-annotation strategy



- DNA sequences from Genbank
- Automatic re-annotation
- Discussion with community (rules, nomenclature, evidences,..)
- Resubmission to Genbank
- Publication

Nomenclature

- Existing references are retained
- Same locus_tag:
 - Genes with unchanged coordinates
 - Genes with changed start coordinate (new protein ID in DB)
- „In between“ locus_tag:
 - Newly annotated genes (e.g. CT_444.1)
- Removed genes lose locus_tag
- Evidence (PMID) of functional annotation is added as note
- Gene names?

References - now

LOCUS AE001363 1230230 bp DNA circular BCT 05-MAR-2010
DEFINITION Chlamydomphila pneumoniae CWL029, complete genome.
ACCESSION AE001363 AE001585-AE001687
VERSION AE001363.1 GI:6626250
DBLINK BioProject: PRJNA248
SOURCE Chlamydomphila pneumoniae CWL029
ORGANISM Chlamydomphila pneumoniae CWL029
Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae;
Chlamydia/Chlamydomphila group; Chlamydia.

REFERENCE 1 (bases 1 to 1230230)
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
PUBMED 10192388

Genome paper

REFERENCE 2 (bases 1 to 1230230)
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L.,
Grimwood,J., Davis,R.W. and Stephens,R.S.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA

Direct submission

References – after re-annotation

LOCUS AE001363 1230230 bp DNA circular BCT 05-MAR-2010
DEFINITION Chlamydomphila pneumoniae CWL029, complete genome.
ACCESSION AE001363 AE001585-AE001687
VERSION AE001363.1 GI:6626250
DBLINK BioProject: PRJNA248
SOURCE Chlamydomphila pneumoniae CWL029
ORGANISM Chlamydomphila pneumoniae CWL029
Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae;
Chlamydia/Chlamydomphila group; Chlamydia.

REFERENCE 1 (bases 1 to 1230230)
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
PUBMED 10192388

Genome paper

REFERENCE 2 (bases 1 to 1230230)
AUTHORS The Chlamydia re-annotation project consortium
TITLE Reannotation of all publicly available chlamydial genomes
JOURNAL XXX (2013)
PUBMED XXX

Re-annotation paper

REFERENCE 3 (bases 1 to 1230230)
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L.,
Grimwood,J., Davis,R.W. and Stephens,R.S.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA

Direct submission



Session schedule

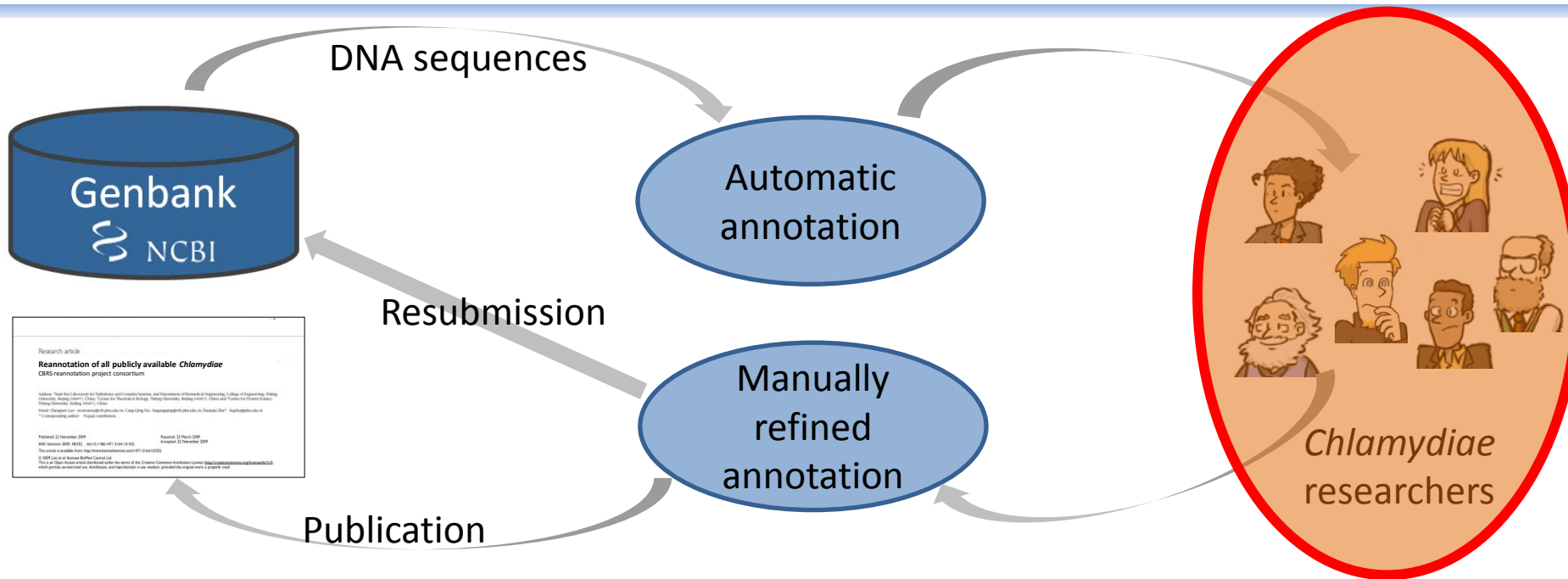
- Intro and automatic annotation (T. Weinmaier)
 - Discussion
- Nomenclature
 - **Discussion**
- Manual refinement
 - Discussion
- Submission / Publication
 - Discussion



Session schedule

- Intro and automatic annotation (T. Weinmaier)
 - Discussion
- Nomenclature
 - Discussion
- **Manual refinement**
 - Discussion
- Submission / Publication
 - Discussion

Proposed re-annotation strategy



- DNA sequences from Genbank
- Automatic re-annotation
- Discussion with community (rules, nomenclature, evidences,..)
- Resubmission to Genbank
- Publication

Manual refinement: your contributions

Locus_tag	Product	Gene name	EC number	Pubmed ID	Taxonomic transfer level
CPn_0008	HB2 protein	hb2		22937701	Chlamydophila pneumoniae
CPn_0056	Phosphoglucomutase/ phosphomannomutase		5.4.2.2	10839753	Chlamydophila pneumoniae
CPn_0663.1	Tryptophanyl tRNA Synthetase	trpS_1		23404517	Chlamydiaeae
CPn1016	Chlamydial protease like activity factor	cPAF		11304554	Chlamydiales
Deleted: CPn1119	Hypothetical protein			10748823	Chlamydophila pneumoniae CWL029

Manual refinement: your contributions

Locus_tag	Product	Gene name	EC number	Pubmed ID	Taxonomic transfer level
CPn_0008	HB2 protein	hb2		22937701	Chlamydophila pneumoniae
CPn_0056	Phosphoglucomutase/ phosphomannomutase		5.4.2.2	10839753	Chlamydophila pneumoniae
CPn_0663.1	Tryptophanyl tRNA Synthetase	trpS_1		23404517	Chlamydiaeae
CPn1016	Chlamydial protease like activity factor	cPAF		11304554	Chlamydiales
Deleted: CPn1119	Hypothetical protein			10748823	Chlamydophila pneumoniae CWL029

Genbank entry - now

```
gene      10975..11688
          /locus_tag="CPn_0008"
CDS       10975..11688
          /locus_tag="CPn_0008"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="AAD18166.1"
          /db_xref="GI:4376264"
          /translation="MISGLLFLLVRREVPTVRSEEIPRGVSVTPSEEPALEKAQKEPE
          TTKILDRLPKELDQLDITYIQEVFACLERLKDPKYEDRGLLTEAKEKLRVFDVVEKDMM
          SEFLDIQRVLNEEAYYVEHCQDPLENIAYEIFSSQELRDYYCAGVCGYLPDARADR
          LKRSVKEVMDRFMRVTWKSWEASVMLDHSYGVARELFKKAVGVLEESVYKILFKSYRD
          AFYECEKAKIQRDGRFKWL"
```

Automatic annotation

Genbank entry – after re-annotation

gene 10975..11688
/locus_tag="CPn_0008"

/gene="hb2"

Manual annotation

CDS 10975..11688
/locus_tag="CPn_0008"

/gene="hb2"

Manual annotation

/codon_start=1

/transl_table=11

/product="HB2 protein"

Manual annotation

/protein_id="AAD18166.1"

/db_xref="GI:4376264"

/note="product, gene name derived from literature (PMID 22937701)"

Indication of source

/translation="MISGLLFLLVRREVPTVRSEEIPRGVSVTPSEEPALEKAQKEPE
TKKILDRLPKELDQLDTYIQEVFACLERLKDPKYEDRGLLTEAKEKLRVFDVVEKDMM
SEFLDIQRVLNEEAYYVEHCQDPLENIAYEIFSSQELRDYYCAGVCGYLPDARADR
LKRSVKEVMDRFMRVTWKSWEASVMLDHSYGVARELFKKAVGVLEESVYKILFKSYRD
AFYECEKAKIQRDGRFKWL"

Genbank entry - now

gene 70953..72749

`/gene="mrsA"`

Automatic annotation

`/locus_tag="CPn_0056"`

CDS 70953..72749

`/gene="mrsA"`

Automatic annotation

`/locus_tag="CPn_0056"`

`/codon_start=1`

`/transl_table=11`

`/product="Phosphomannomutase"`

Automatic annotation

`/protein_id="AAD18209.1"`

`/db_xref="GI:4376312"`

`/translation="MKEVEQRIRSLYDAVTAENICRWLSNDCTQQDAKTILGWLDTP`

`AQLEDLFGATLTFGTGGLRSLMGIGTNRINLFTIRRTTQGLVQVLR AHLPHPGDPMRV`

....

Genbank entry – after re-annotation

gene 70953..72749

`/gene="mrsA"`

Automatic annotation

`/locus_tag="CPn_0056"`

CDS 70953..72749

`/gene="mrsA"`

Automatic annotation

`/locus_tag="CPn_0056"`

`/codon_start=1`

`/transl_table=11`

`/product="Phosphoglucomutase/phosphomannomutase"`

Manual annotation

`/EC_number="5.4.2.2"`

Manual annotation

`/protein_id="AAD18209.1"`

`/db_xref="GI:4376312"`

`/note="product, EC_number derived from literature (PMID 10839753)"`

Indication of source

`/translation="MKEVEQRIRSLYDAVTAENICRWLSNDCTQQDAKTILGWLDTDP`

`AQLEDLFGATLTFGTGGLRSLMGIGTNRINLFTIRRTTQGLVQVLR AHLPHPGDPMRV`

....



Session schedule

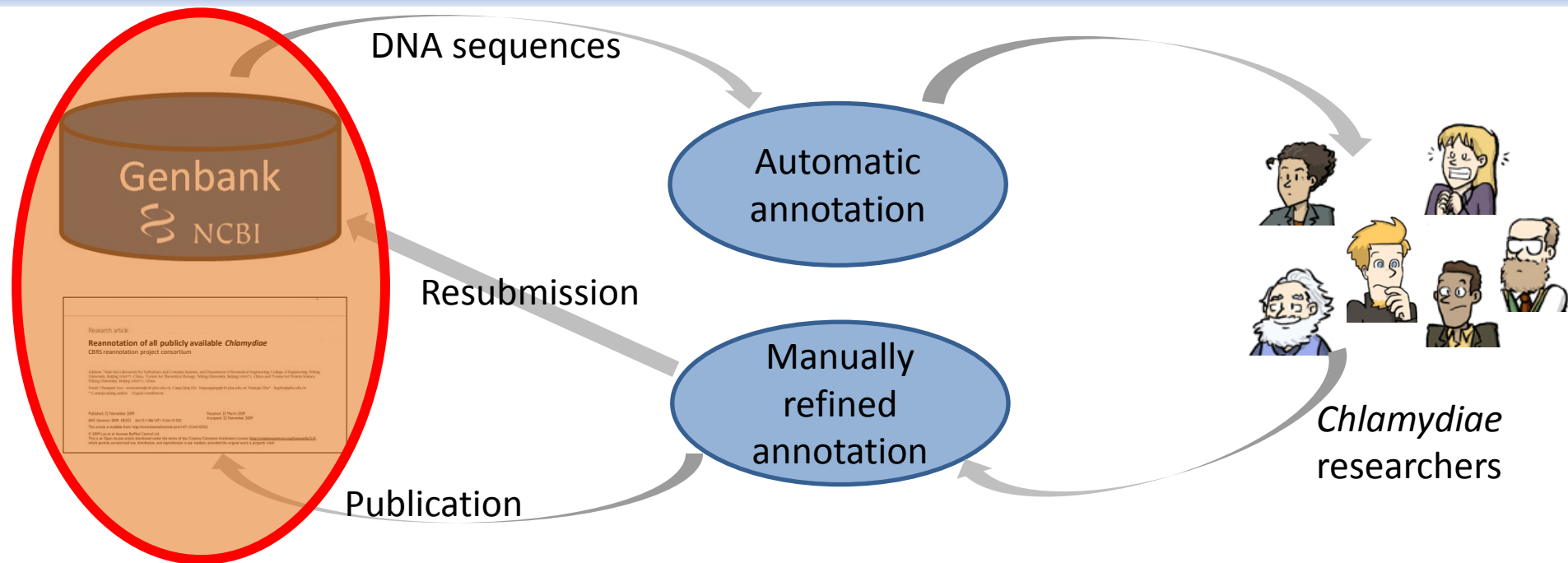
- Intro and automatic annotation (T. Weinmaier)
 - Discussion
- Nomenclature
 - Discussion
- Manual refinement
 - **Discussion**
- Submission / Publication
 - Discussion



Session schedule

- Intro and automatic annotation (T. Weinmaier)
 - Discussion
- Nomenclature
 - Discussion
- Manual refinement
 - Discussion
- **Submission / Publication**
 - Discussion

Proposed re-annotation strategy



- DNA sequences from Genbank
- Automatic re-annotation
- Discussion with community (rules, nomenclature, evidences,..)
- Resubmission to Genbank
- Publication

Submission

Why?

- Provides up-to-date knowledge for BLAST searches
- Facilitates future (multi-)genome projects

How?

- Chlamydia re-annotation project generates new Genbank files
- Owners of Genbank entries re-submit

Publication

- Authors:
all contributors to manual re-annotation
- Manuscript outline
 - Goal: improving existing genome annotations
 - Method: community-based re-annotation
 - Results
 - Comparison between old and new annotation
 - Analysis of the chlamydial pan-genome
- Further ideas / contributions?

Research article

Re-annotation of all publicly available chlamydial genomes CBRS reannotation project consortium

Address: State Key Laboratory for Infection and Complex Systems, and Department of Biomedical Engineering, College of Engineering, Peking University, Beijing 100871, China; Center for Biomedical Biology, Peking University, Beijing 100871, China and Center for Protein Science, Peking University, Beijing 100871, China

Email: Changyan Liu: liuchangyan@pku.edu.cn; Gang-Qing He: hgqiang@pku.edu.cn; Guangjin Shi* shiguangjin@pku.edu.cn

* Corresponding author. shiguangjin@pku.edu.cn

Published: 22 November 2009

BMC Genomics 2009, 10:522

doi:10.1186/1471-2164-10-522

This article is available from: <http://www.biomedcentral.com/1471-2164/10/522>

© 2009 Liu et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 22 March 2009

Accepted: 22 November 2009



Session schedule

- Intro and automatic annotation (T. Weinmaier)
 - Discussion
- Nomenclature
 - Discussion
- Manual refinement
 - Discussion
- Submission / Publication
 - **Discussion**

Thank you for your attention!

