Moving on with Standard interoperability For modelling in biology

The “WorldWide Web consortium” of modelling in biology

http://co.mbine.org/

2012 Jul 20

Cell

A Whole-Cell Computational Model Predicts Phenotype from Genotype

Jonathan R. Karr,[1,4] Jayodita C. Sanghvi,[2,4] Derek N. Macklin,[2] Miriam V. Gutschow,[2] Jared M. Jacobs,[2] Benjamin Bolival, Jr.,[2] Nacyra Assad-Garcia,[3] John I. Glass,[3] and Markus W. Covert[2,*]

[1]Graduate Program in Biophysics
[2]Department of Bioengineering
Stanford University, Stanford, CA 94305, USA
[3]J. Craig Venter Institute, Rockville, MD 20850, USA
[4]These authors contributed equally to this work
*Correspondence: mcovert@stanford.edu
http://dx.doi.org/10.1016/j.cell.2012.05.044

SUMMARY

Understanding how complex phenotypes arise from individual molecules and their interactions is a primary challenge in biology that computational approaches are poised to tackle. We report a whole-cell computational model of the life cycle of the human pathogen *Mycoplasma genitalium* that includes all of its molecular components and their interactions. An integrative approach to modeling that combines diverse mathematics enabled the simultaneous inclusion of fundamentally different cellular processes and experimental measurements. Our whole-cell model accounts for all annotated gene functions and was validated against a broad range of data. The model provides insights into many previously unobserved cellular behaviors, including in vivo rates of protein-DNA association

First, until recently, not enough has been known about the individual molecules and their interactions to completely model any one organism. The advent of genomics and other high-throughput measurement techniques has accelerated the characterization of some organisms to the extent that comprehensive modeling is now possible. For example, the mycoplasmas, a genus of bacteria with relatively small genomes that includes several pathogens, have recently been the subject of an exhaustive experimental effort by a European consortium to determine the transcriptome (Güell et al., 2009), proteome (Kühner et al., 2009), and metabolome (Yus et al., 2009) of these organisms.

The second limiting factor has been that no single computational method is sufficient to explain complex phenotypes in terms of molecular components and their interactions. The first approaches to modeling cellular physiology, based on ordinary differential equations (ODEs) (Atlas et al., 2008; Browning et al., 2004; Castellanos et al., 2004, 2007; Domach et al., 1984; Tomita et al., 1999), were limited by the difficulty in obtaining the necessary model parameters. Subsequently, alternative

Cell

A Whole-Cell Computational Model Predicts Phenotype from Genotype

Jonathan R. Karr,[1,4] Jayodita C. Sanghvi,[2,4] Derek N. Macklin,[2] Miriam V. Gutschow,[2] Jared M. Jacobs,[2] Benjamin Bolival, Jr.,[2] Nacyra Assad-Garcia,[3] John I. Glass,[3] and Markus W. Covert[2,*]

[1]Graduate Program in Biophysics
[2]Department of Bioengineering
Stanford University, Stanford, CA 94305, USA
[3]J. Craig Venter Institute, Rockville, MD 20850, USA
[4]These authors contributed equally to this work
*Correspondence: mcovert@stanford.edu
http://dx.doi.org/10.1016/j.cell.2012.05.044

SUMMARY

Understanding how complex phenotypes arise from individual molecules and their interactions is a primary challenge in biology that computational approaches are poised to tackle. We report a whole-cell computational model of the life cycle of the human pathogen *Mycoplasma genitalium* that includes all of its molecular components and their interactions. An integrative approach to modeling that combines diverse mathematics enabled the simultaneous inclusion of fundamentally different cellular processes and experimental measurements. Our whole-cell model accounts for all annotated gene functions and was validated against a broad range of data. The model provides insights into many previously unobserved cellular behaviors, including in vivo rates of protein-DNA association

First, until recently, not enough has been known about the individual molecules and their interactions to completely model any one organism. The advent of genomics and other high-throughput measurement techniques has accelerated the characterization of some organisms to the extent that comprehensive modeling is now possible. For example, the mycoplasmas, a genus of bacteria with relatively small genomes that includes several pathogens, have recently been the subject of an exhaustive experimental effort by a European consortium to determine the transcriptome (Güell et al., 2009), proteome (Kühner et al., 2009), and metabolome (Yus et al., 2009) of these organisms.

The second limiting factor has been that no single computational method is sufficient to explain complex phenotypes in terms of molecular components and their interactions. The first approaches to modeling cellular physiology, based on ordinary differential equations (ODEs) (Atlas et al., 2008; Browning et al., 2004; Castellanos et al., 2004, 2007; Domach et al., 1984; Tomita et al., 1999), were limited by the difficulty in obtaining the necessary model parameters. Subsequently, alternative

Cell

A Whole-Cell Computational Model Predicts Phenotype from Genotype

Jonathan R. Karr,[1,4] Jayodita C. Sanghvi,[2,4] Derek N. Macklin,[2] Miriam V. Gutschow,[2] Jared M. Jacobs,[2] Benjamin Bolival, Jr.,[2] Nacyra Assad-Garcia,[3] John I. Glass,[3] and Markus W. Covert[2,*]

[1]Graduate Program in Biophysics
[2]Department of Bioengineering
Stanford University, Stanford, CA 94305, USA
[3]J. Craig Venter Institute, Rockville, MD 20850, USA
[4]These authors contributed equally to this work
*Correspondence: mcovert@stanford.edu
http://dx.doi.org/10.1016/j.cell.2012.05.044

SUMMARY

Understanding how complex phenotypes arise from individual molecules and their interactions is a primary challenge in biology that computational approaches are poised to tackle. We report a whole-cell computational model of the life cycle of the human pathogen *Mycoplasma genitalium* that includes all of its molecular components and their interactions. An integrative approach to modeling that combines diverse mathematics enabled the simultaneous inclusion of fundamentally different cellular processes and experimental measurements. Our whole-cell model accounts for all annotated gene functions and was validated against a broad range of data. The model provides insights into many previously unobserved cellular behaviors, including in vivo rates of protein-DNA association

First, until recently, not enough has been known about the individual molecules and their interactions to completely model any one organism. The advent of genomics and other high-throughput measurement techniques has accelerated the characterization of some organisms to the extent that comprehensive modeling is now possible. For example, the mycoplasmas, a genus of bacteria with relatively small genomes that includes several pathogens, have recently been the subject of an exhaustive experimental effort by a European consortium to determine the transcriptome (Güell et al., 2009), proteome (Kühner et al., 2009), and metabolome (Yus et al., 2009) of these organisms.

The second limiting factor has been that no single computational method is sufficient to explain complex phenotypes in terms of molecular components and their interactions. The first approaches to modeling cellular physiology, based on ordinary differential equations (ODEs) (Atlas et al., 2008; Browning et al., 2004; Castellanos et al., 2004, 2007; Domach et al., 1984; Tomita et al., 1999), were limited by the difficulty in obtaining the necessary model parameters. Subsequently, alternative

Cell

A Whole-Cell Computational Model Predicts Phenotype from Genotype

Jonathan R. Karr,[1,4] Jayodita C. Sanghvi,[2,4] Derek N. Macklin,[2] Miriam V. Gutschow,[2] Jared M. Jacobs,[2] Benjamin Bolival, Jr.,[2] Nacyra Assad-Garcia,[3] John I. Glass,[3] and Markus W. Covert[2,*]

[1]Graduate Program in Biophysics
[2]Department of Bioengineering
Stanford University, Stanford, CA 94305, USA
[3]J. Craig Venter Institute, Rockville, MD 20850, USA
[4]These authors contributed equally to this work
*Correspondence: mcovert@stanford.edu
http://dx.doi.org/10.1016/j.cell.2012.05.044

SUMMARY

Understanding how complex phenotypes arise from individual molecules and their interactions is a primary challenge in biology that computational approaches are poised to tackle. We report a whole-cell computational model of the life cycle of the human pathogen *Mycoplasma genitalium* that includes all of its molecular components and their interactions. An integrative approach to modeling that combines diverse mathematics enabled the simultaneous inclusion of fundamentally different cellular processes and experimental measurements. Our whole-cell model accounts for all annotated gene functions and was validated against a broad range of data. The model provides insights into many previously unobserved cellular behaviors, including in vivo rates of protein-DNA association

First, until recently, not enough has been known about the individual molecules and their interactions to completely model any one organism. The advent of genomics and other high-throughput measurement techniques has accelerated the characterization of some organisms to the extent that comprehensive modeling is now possible. For example, the mycoplasmas, a genus of bacteria with relatively small genomes that includes several pathogens, have recently been the subject of an exhaustive experimental effort by a European consortium to determine the transcriptome (Güell et al., 2009), proteome (Kühner et al., 2009), and metabolome (Yus et al., 2009) of these organisms.

The second limiting factor has been that no single computational method is sufficient to explain complex phenotypes in terms of molecular components and their interactions. The first approaches to modeling cellular physiology, based on ordinary differential equations (ODEs) (Atlas et al., 2008; Browning et al., 2004; Castellanos et al., 2004, 2007; Domach et al., 1984; Tomita et al., 1999), were limited by the difficulty in obtaining the necessary model parameters. Subsequently, alternative

Theory

Cell

A Whole-Cell Computational Model Predicts Phenotype from Genotype

Jonathan R. Karr,[1,4] Jayodita C. Sanghvi,[2,4] Derek N. Macklin,[2] Miriam V. Gutschow,[2] Jared M. Jacobs,[2] Benjamin Bolival, Jr.,[2] Nacyra Assad-Garcia,[3] John I. Glass,[3] and Markus W. Covert[2,*]

[1]Graduate Program in Biophysics
[2]Department of Bioengineering
Stanford University, Stanford, CA 94305, USA
[3]J. Craig Venter Institute, Rockville, MD 20850, USA
[4]These authors contributed equally to this work
*Correspondence: mcovert@stanford.edu
http://dx.doi.org/10.1016/j.cell.2012.05.044

SUMMARY

Understanding how complex phenotypes arise from individual molecules and their interactions is a primary challenge in biology that computational approaches are poised to tackle. We report a whole-cell computational model of the life cycle of the human pathogen *Mycoplasma genitalium* that includes all of its molecular components and their interactions. An integrative approach to modeling that combines diverse mathematics enabled the simultaneous inclusion of fundamentally different cellular processes and experimental measurements. Our whole-cell model accounts for all annotated gene functions and was validated against a broad range of data. The model provides insights into many previously unobserved cellular behaviors, including in vivo rates of protein-DNA association

First, until recently, not enough has been known about the individual molecules and their interactions to completely model any one organism. The advent of genomics and other high-throughput measurement techniques has accelerated the characterization of some organisms to the extent that comprehensive modeling is now possible. For example, the mycoplasmas, a genus of bacteria with relatively small genomes that includes several pathogens, have recently been the subject of an exhaustive experimental effort by a European consortium to determine the transcriptome (Güell et al., 2009), proteome (Kühner et al., 2009), and metabolome (Yus et al., 2009) of these organisms.

The second limiting factor has been that no single computational method is sufficient to explain complex phenotypes in terms of molecular components and their interactions. The first approaches to modeling cellular physiology, based on ordinary differential equations (ODEs) (Atlas et al., 2008; Browning et al., 2004; Castellanos et al., 2004, 2007; Domach et al., 1984; Tomita et al., 1999), were limited by the difficulty in obtaining the necessary model parameters. Subsequently, alternative

Cell

A Whole-Cell Computational Model Predicts Phenotype from Genotype

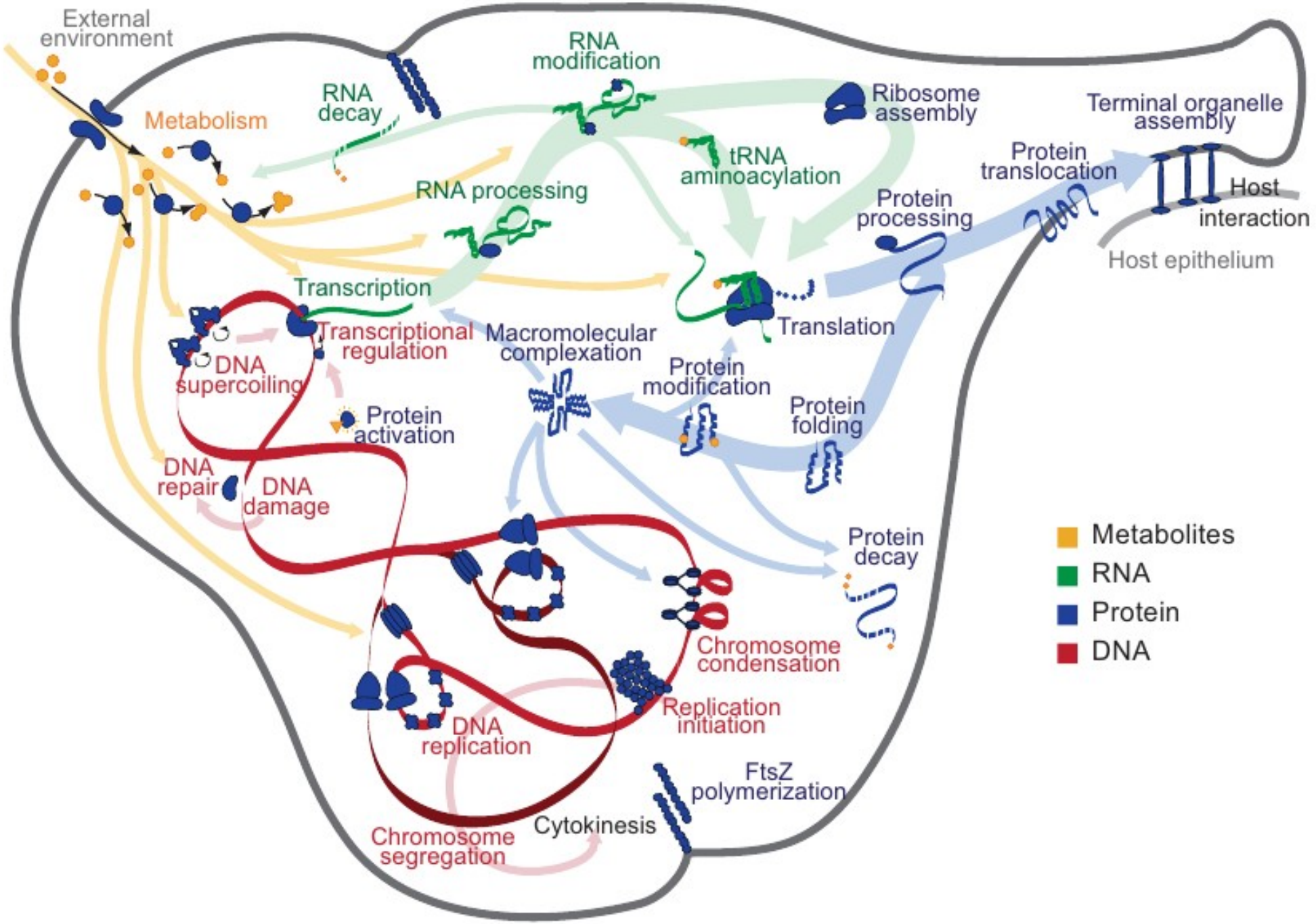
Jonathan R. Karr,[1,4] Jayodita C. Sanghvi,[2,4] Derek N. Macklin,[2] Miriam V. Gutschow,[3] Jared M. Jacobs,[2] Benjamin Bolival, Jr.,[2] Nacyra Assad-Garcia,[3] John I. Glass,[3] and Markus W. Covert[2,*]
[1]Graduate Program in Biophysics
[2]Department of Bioengineering
Stanford University, Stanford, CA 94305, USA
[3]J. Craig Venter Institute, Rockville, MD 20850, USA
[4]These authors contributed equally to this work
*Correspondence: mcovert@stanford.edu
http://dx.doi.org/10.1016/j.cell.2012.05.044

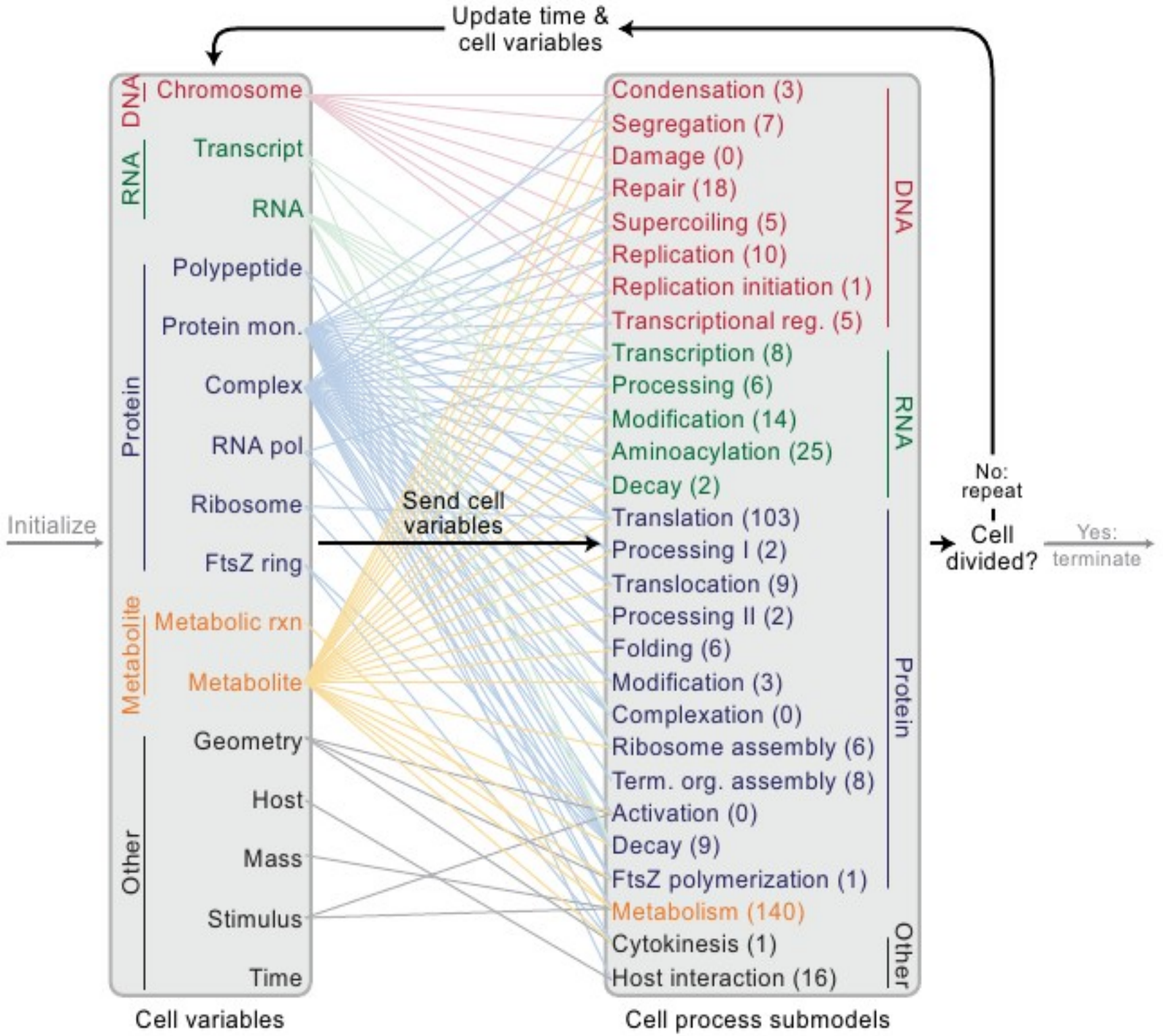
SUMMARY

Understanding how complex phenotypes arise from individual molecules and their interactions is a primary challenge in biology that computational approaches are poised to tackle. We report a whole-cell computational model of the life cycle of the human pathogen *Mycoplasma genitalium* that includes all of its molecular components and their interactions. An integrative approach to modeling that combines diverse mathematics enabled the simultaneous inclusion of fundamentally different cellular processes and experimental measurements. Our whole-cell model accounts for all annotated gene functions and was validated against a broad range of data. The model provides insights into many previously unobserved cellular behaviors, including in vivo rates of protein-DNA association

First, until recently, not enough has been known about the individual molecules and their interactions to completely model any one organism. The advent of genomics and other high-throughput measurement techniques has accelerated the characterization of some organisms to the extent that comprehensive modeling is now possible. For example, the mycoplasmas, a genus of bacteria with relatively small genomes that includes several pathogens, have recently been the subject of an exhaustive experimental effort by a European consortium to determine the transcriptome (Güell et al., 2009), proteome (Kühner et al., 2009), and metabolome (Yus et al., 2009) of these organisms.

The second limiting factor has been that no single computational method is sufficient to explain complex phenotypes in terms of molecular components and their interactions. The first approaches to modeling cellular physiology, based on ordinary differential equations (ODEs) (Atlas et al., 2008; Browning et al., 2004; Castellanos et al., 2004, 2007; Domach et al., 1984; Tomita et al., 1999), were limited by the difficulty in obtaining the necessary model parameters. Subsequently, alternative


External environment
Metabolism
RNA decay
RNA modification
Ribosome assembly
tRNA aminoacylation
RNA processing
Terminal organelle assembly
Protein translocation
Host interaction
Host epithelium
Protein processing
Transcription
Transcriptional regulation
DNA supercoiling
Macromolecular complexation
Translation
Protein modification
Protein activation
Protein folding
DNA repair
DNA damage
Protein decay
Chromosome condensation
Replication initiation
DNA replication
FtsZ polymerization
Cytokinesis
Chromosome segregation
Metabolites
RNA
Protein
DNA


Update time & cell variables
DNA
Chromosome
RNA
Transcript
RNA
Protein
Polypeptide
Protein mon.
Complex
RNA pol
Ribosome
FtsZ ring
Metabolite
Metabolic rxn
Metabolite
Other
Geometry
Host
Mass
Stimulus
Time
Initialize
Send cell variables
Condensation (3)
Segregation (7)
Damage (0)
Repair (18)
Supercoiling (5)
Replication (10)
Replication initiation (1)
Transcriptional reg. (5)
Transcription (8)
Processing (6)
Modification (14)
Aminoacylation (25)
Decay (2)
Translation (103)
Processing I (2)
Translocation (9)
Processing II (2)
Folding (6)
Modification (3)
Complexation (0)
Ribosome assembly (6)
Term. org. assembly (8)
Activation (0)
Decay (9)
FtsZ polymerization (1)
Metabolism (140)
Cytokinesis (1)
Host interaction (16)
DNA
RNA
Protein
Other
No: repeat
Cell divided?
Yes: terminate
Cell variables
Cell process submodels

Do they encode their model with standards?

No ... But they could have!

Why did they not?

The model is a set of 28 modules, each modelled with a specific approach. Most of the approaches are covered by SBML L3, some are not.

The modules are connected at simulation time. This could be encoded with the forthcoming SED-ML nested extension.

But no software could run the model ... MatLab can. This is a general threat for open tools in computational systems biology. Cytoscape, SBW, Garuda etc.?

They are (excellent) biologists. They addressed their questions with the most efficient method. They are not the bad guys!

The journal Cell does not request deposition in standard format, therefore, no counter-pressure.

Why COMBINE?

(slide written in March 2010)

- Current efforts are almost entirely dependent on key people (SBML: Mike Hucka, CellML: Peter Hunter/Poul Nielsen, NeuroML: Padraig Gleeson, SBGN: NLN, BioPAX: Emek Demir/Gary Bader). Their disengagement means disaggregation.
- Current funding structure is fragile. Many different grants, sometimes only supporting meetings (SBGN), none of them infrastructure rolling funding, often tied to individuals.
- Current efforts are not immune against intellectual property claims that would destroy the community (e.g. Caltech and SBML)
- Existing standards are developed with very different approaches, quality checks, and are based on completely different assumptions (e.g. NeuroML assumes implicit knowledge embedded in simulators, SBML explicitely describe all mathematics)
- APIs needs industry-grade support, incompatible with standard academic usages and possibilities

So many meetings ...

BioPAX face 2 face

SBML forum

SBGN meeting

SBML hackathon

BioModels training camp

SuperHackathon

CellML workshop

NeuroML workshop

Multiple involvements

Authors on several publications describing “our” standards:

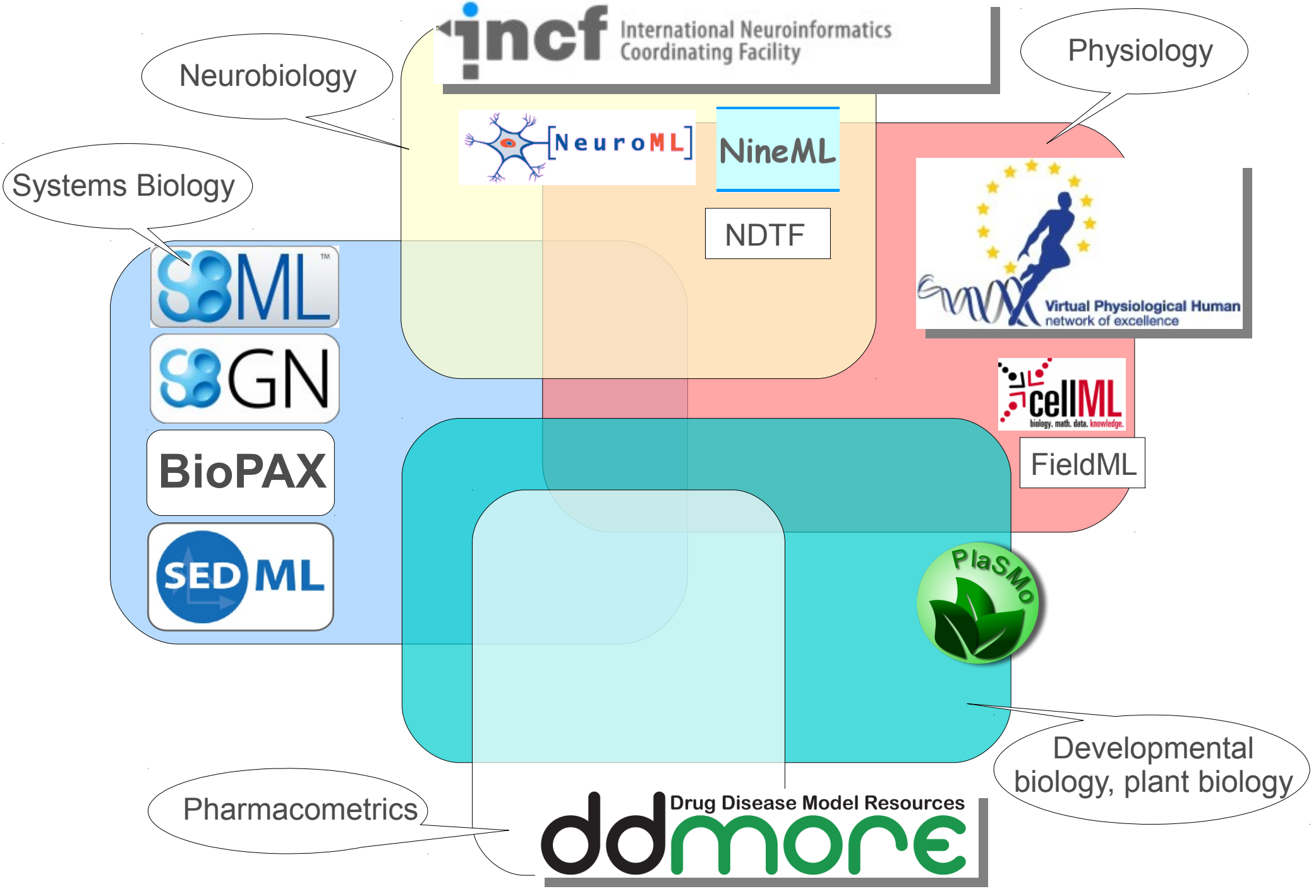
Mirit Aladjem (BioPAX, SBGN), *Frank Bergmann* (SBGN, SED-ML), *Emek Demir* (BioPAX, SBGN), *Mélanie Courtot* (SBGN, SBO/KiSAO), *Andrew Finney* (BioPAX, SBML, SBO/KiSAO), *Igor Goryanin* (SBGN, SBML), *Stefan Hoops* (SBO/KiSAO, SED-ML), *Michael Hucka* (BioPAX, SBGN, SBML, SBO/KiSAO, SED-ML), *Peter Hunter* (CellML, SBML), *Nick Juty* (SBML, SBO/KiSAO), *Douglas Kell* (SBGN, SBO/KiSAO), *Hiroaki Kitano* (SBML, SBGN), *Fedor Kolpakov* (SBGN, SED-ML), *Nicolas Le Novère* (BioPAX, SBGN, SBML, SBO/KiSAO, SED-ML), *Pedro Mendes* (SBO/KiSAO), Huaiyu Mi (BioPAX, SBGN), *David Nickerson* (CellML, SED-ML), *Poul Nielsen* (CellML, SBML), *Sven Sahle* (SBGN, SED-ML), *Herbert Sauro* (SBGN, SBML), *Jacky Snoep* (SBGN, SBO/KiSAO), *Alice Villéger* (SBGN, SBO/KiSAO), *Dagmar Waltemath* (SED-ML, SBO/KiSAO), *Sarala Wimalaratne* (BioPAX, SBGN, SBO/KiSAO)

Mission 1: Coordinating meetings

- Annual COMBINE forums
  - COMBINE 2010: October 6–9, Edinburgh, **81 attendees**
  - COMBINE 2011: September 3-7, Heidelberg, **89 attendees**
  - COMBINE 2012: August 15-19, Toronto
  - COMBINE 2013: Date unknown, location unknown
    http://www.surveymonkey.com/s/combine-harmony-hosting-interest

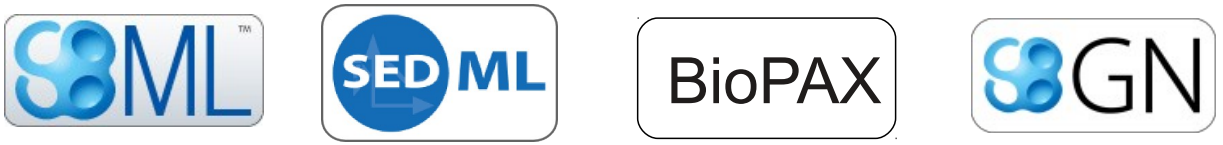
- The Hackathons on Resources for Modeling in Biology
  - HARMONY 2011: April 18-22, New-York City, **59 attendees**
  - HARMONY 2012; May 21-25, Maastricht, **60 attendees**
  - HARMONY 2013: Spring 2013, UCHC, USA
  - HARMONY 2014: Date unknown, location unknown
    http://www.surveymonkey.com/s/combine-harmony-hosting-interest

Parallel and redundant efforts



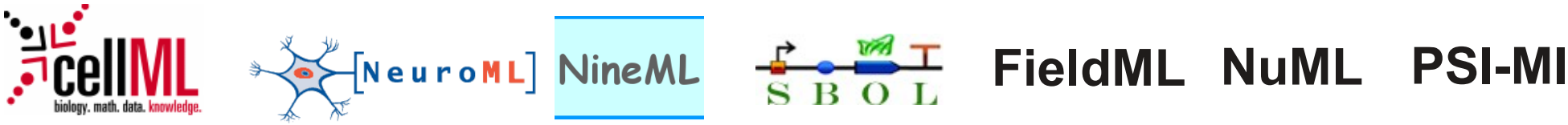
Mission 2: Coordinating standards development

- **CORE STANDARDS:** Efforts fulfilling COMBINE criteria and aiming at following COMBINE rules and interoperate with other COMBINE standards



- **ASSOCIATED EFFORTS:** Standards that are not representation formats, but aiming at enrich or bridge the core standards

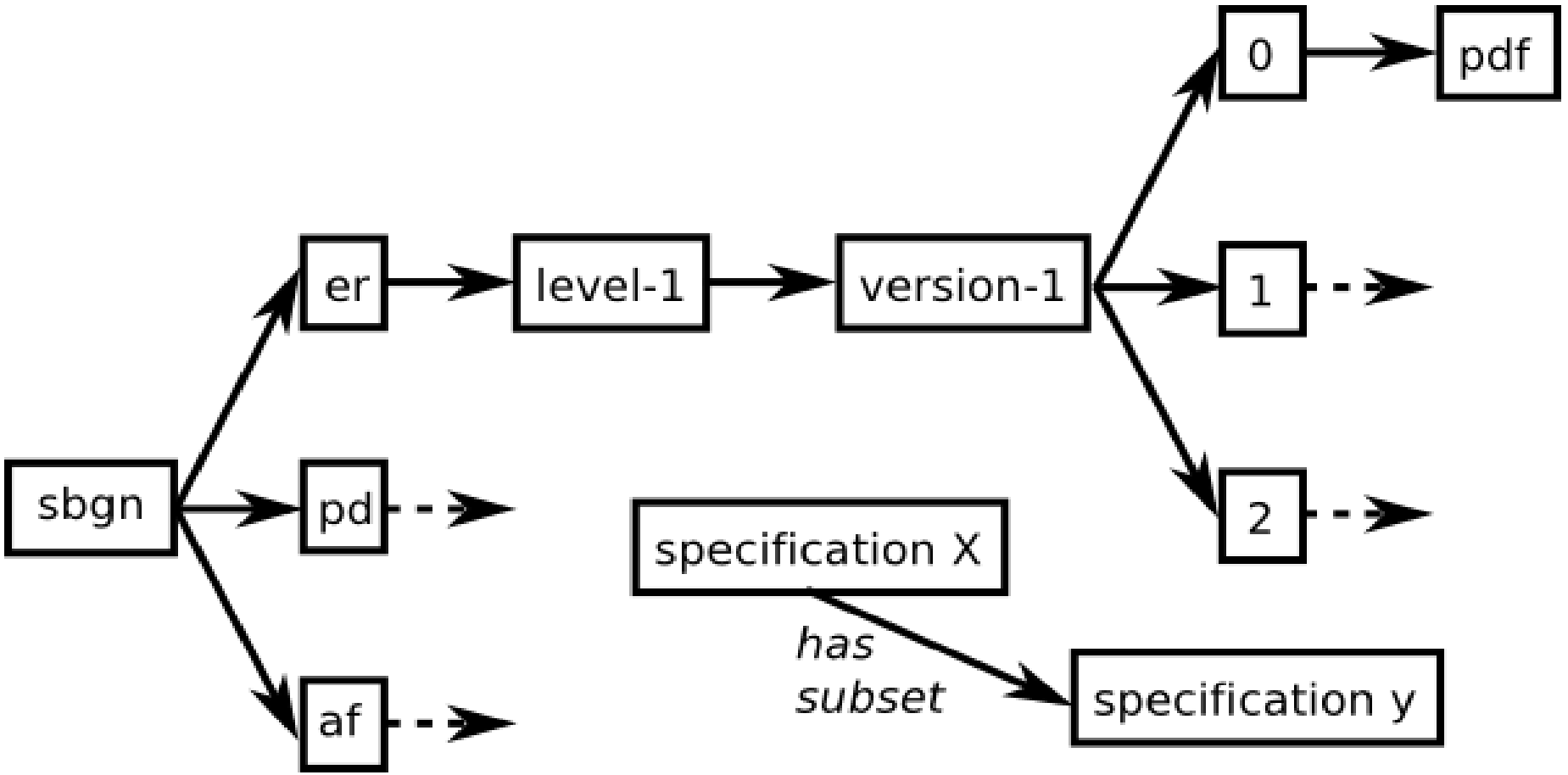
- **RELATED EFFORTS:** Formats developed by other communities, that complement or interoperate with COMBINE formats, and that we would like to see joining COMBINE or collaborating closely to COMBINE



Mission 3: Developing SOPs and common tools

- Technical requirements
  - Common metadata structures and vocabularies – predate COMBINE
  - Infrastructure to support format specifications
  - Archive format to bundle COMBINE formats together
  - ...
- Governance
  - Criteria to meet in order to join COMBINE core formats - started
  - How to organise elections for editorships - on the way
  - How to organise COMBINE meetings (forums or HARMONY) – started
  - ...

Infrastructure for specification documents



Retrieval of specification documents

http://identifiers.org/combine.specifications/sbgn
http://identifiers.org/combine.specifications/sbgn.er
http://identifiers.org/combine.specifications/sbgn.er.level-1
http://identifiers.org/combine.specifications/sbgn.er.level-1.version-1
http://identifiers.org/combine.specifications/sbgn.er.level-1.version-1.0
http://identifiers.org/combine.specifications/sbgn.er.level-1.version-1.0.pdf

redirects

This never changes

http://co.mbine.org/specifications/sbgn
http://co.mbine.org/specifications/sbgn.er
http://co.mbine.org/specifications/sbgn.er.level-1
http://co.mbine.org/specifications/sbgn.er.level-1.version-1
http://co.mbine.org/specifications/sbgn.er.level-1.version-1.0
http://co.mbine.org/specifications/sbgn.er.level-1.version-1.0.pdf

alias

This can change over time

http://co.mbine.org/standards/sbgn
http://co.mbine.org/standards/sbgn.er


http://co.mbine.org/standards/sbgn/er/level-1/version-1/2
http://co.mbine.org/standards/sbgn/er/level-1/version-1/2
http://co.mbine.org/standards/sbgn/er/level-1/version-1/0


sbgn.er.level-1.version-1.0.pdf

Systems Biology Graphical Notation

View | Edit | Revisions | Access control

The Systems Biology Graphical Notation (SBGN), is a set standard graphical languages to describe biological knowledge. It is currently made up of three languages describing Process Descriptions, Entity Relationships and Activity Flows.

Normative definitions

SBGN is defined by the a set of specification documents, that define the symbols used in the languages, and the rules to assemble them in maps. The latest specifications are:

- SBGN PD Level 1 Version 1.3
- SBGN ER Level 1 Version 1.2
- SBGN AF Level 1 Version 1.0

http://co.mbine.org/standards/sbgn

Governance

SBGN development is coordinated by an editorial board elected by the community, and a scientific committee made up of PIs of SBGN supporting grants and invited members.

Communication

SBGN development is discussed through mailing-list, the main one being sbgn-discuss@caltech.edu.

Software support

Several data resources and software claim support for SBGN. This includes an API is also available to help implementing support: libSBGN.

Contact

http://co.mbine.org/standards/sbgn/er

Home › SBGN Entity Relationships

SBGN Entity Relationships

View | Edit | Access control

The SBGN Entity Relationship (ER) language allows you to see all the relationships in which a given entity participates, regardless of the temporal aspects. Relationships can be seen as rules describing the influences of entities nodes on other relationships.

The last specifications of SBGN ER is SBGN ER Level 1 Version 1.2.

http://co.mbine.org/standards/sbgn/er/level-1/version-1/2

Home › SBGN ER Level 1 Version 1.2

SBGN ER Level 1 Version 1.2

View | Edit | Revisions | Access control

Version 1.2 of Level 1 of the SBGN Entity Relationship Language was published on 14 April 2011.

The specification can be found at:

- http://co.mbine.org/specifications/sbgn.er.level-1.version-1.2.pdf
- http://dx.doi.org/10.1038/npre.2011.5902.1
- http://sbgn.svn.sourceforge.net/viewvc/sbgn/EntityRelationship/tags/Level1-Version1.2/sbgn_ER-level1.pdf

Identifier for this specification is: http://identifiers.org/combine.specifications/sbgn.er.level-1.version-1.2

To cite this document, please use:

Nicolas Le Novère, Emek Demir, Huaiyu Mi, Stuart Moodie, Alice Villéger. Systems Biology Graphical Notation: Entity Relationship language Level 1, Version 1.2. Available from COMBINE <http://identifiers.org/combine.specifications/sbgn.er.level-1.version-1.2> (2011)

http://co.mbine.org/standards/sbgn/er/level-1/version-1/2

Home › SBGN ER Level 1 Version 1.2

SBGN ER Level 1 Version 1.2

View | Edit | Revisions | Access control

Version 1.2 of Level 1 of the SBGN Entity Relationship Language was published on 14 April 2011.

The specification can be found at:

- http://co.mbine.org/specifications/sbgn.er.level-1.version-1.2.pdf
- http://dx.doi.org/10.1038/npre.2011.5902.1
- http://sbgn.svn.sourceforge.net/viewvc/sbgn/EntityRelationship/tags/Level1-Version1.2/sbgn_ER-level1.pdf

Identifier for this specification is: http://identifiers.org/combine.specifications/sbgn.er.level-1.version-1.2

To cite this document, please use:

Nicolas Le Novère, Emek Demir, Huaiyu Mi, Stuart Moodie, Alice Villéger. Systems Biology Graphical Notation: Entity Relationship language Level 1, Version 1.2. Available from COMBINE <http://identifiers.org/combine.specifications/sbgn.er.level-1.version-1.2> (2011)

More at http://co.mbine.org/standards/specification-infrastructure

COMBINE archive

- Most software only open 1 file at a time. One downloads 1 file from a repository, or as a supplementary material, but
- The SBML comp package will need to access several files. SED-ML files can refer to several models. NuML will need to be linked to model and simulation descriptions. SBGN map enrichment may need information contained in BioPAX, SBML or NuML files. Other types of models, such as PKPD or physiological, require more than the equations.
- The COMBINE archive is a "zip" file, with the extension .omex, for "Open Modeling EXchange format".
- The archive contains a manifest file that describes the location and the type of each data file contained in the archive, a metadata file containing clerical information about the various files contained in the archive, and the archive itself, and all the remaining files necessary to the model and simulation project.
- Presentation Saturday 18th, 15:30
- More at http://co.mbine.org/documents/archive

Mission 4: Recognised voice

- COMBINE aims to become a “standardisation” body
  - This means a quality label. A “COMBINE standard” is a guarantee of stability, community endorsement, support etc.
  - COMBINE production can be used in SOPs at other organisations
  - COMBINE must be an actor, on par with CSG, FGED, INCF, PSI etc.
- Single point of contact with user organisations including Industry
  - Tool developers (General platforms or specific tools)
  - Publishers
  - Pharmaceutical industry
- A point of contact for funding bodies
- A point of contact for legal entities, e.g. government and regulatory bodies

How did we perform so far?

Mission 1 (meetings)

Mission 2 (format development)

Mission 3 (SOPs and tools)

Mission 4 (Identity)

How did we perform so far?

Mission 1 (meetings)

✔✔

Mission 2 (format development)

✔✘

Still overlap between formats, redundant developments, lack of common structures

Mission 3 (SOPs and tools)

✔✘

Starting on the technical side, almost non-existent on the governance one. Community does not use common tools

Mission 4 (Identity)

✘✘

COMBINE not recognised, either by the community or outside. No visibility. No common funding. No common vision

How do-we move forward?

1) What do we want COMBINE to be?

2) Participation

3) Structure and governance

1) What do we want COMBINE to be?

A set of meetings	✔
A community infrastructure	?
A set of common guidelines	?
A registry of standards	?
A standardisation body	?
A legal entity	?

2) Participation

- How to get firmer endorsement from independent standard efforts? We can hardly afford conditional or intermittent endorsement. We need to scale up selflessness. From *“the good of **ML before the need of my specific project”* to *“the good of COMBINE community before the wishes of **ML”.* There is a cooperative return on confidence investment.
- How to encourage participation? to the discussions, to the documents, to the meetings, to the elections, to the website, to the funding, to the advocacy etc.
  (It is not enough to meet twice a year in nice location, and have a fun time with friends chatting and drinking beer. All that is necessary, but not sufficient.)
- We need to be proactive and recruit participants from the communities of modellers, pathway creators, database developers etc. (the users) Plus seek the support of VIPs in systems biology. One word to the right hear can be very helpful or terribly destructive.

3) Governance

- We should have a governance structure that is:
  - Well defined: described in a document, intelligible, justifiable and communicable
  - Robust: be small enough to function, but big enough to deliver. It should be accountable, with anti-stalling devices
  - Sustainable: it should not depend on a given person or institution or funding
- Many different models exist. We should look around and imitate, without adopting everything.

INCF

- Hosted by Karolinska Institutet and the Royal Institute of Technology
- Funded by member countries (16), proportional to *gross domestic expenditures on research and development (GERD)*
- $2.4 Million/annum
- Secretariat: 13 people
- Executive director
- Governing board: 1/2 per member + observers (EU)
- Coordination activity, but does not “provide” standards, tools or databases

W3C

- 3 Host Institutions: MIT, ERCIM, and Keio University
- Funded by members (373), depending on country, type of organisation and size of income. Can be companies, Universities, research grants, sponsorships
- Management team: 16 people
- W3C staff: 68 people
- 1 director and 1 CEO
- Regional offices
- Advisory committee: 1 per member
- Advisory board, elected by above
- Chartered groups
- Coordination activity, but also provides standards (not by staff)

HUPO-PSI

- Not independent but part of HUPO (what would be the equivalent for COMBINE?)
- 1 steering committee (8 members)
- Interest groups (5 groups)
- Each has chair, co-chairs, format editors, MI lead, ontology lead, secretary
- HUPO:
- Council, executive committee, committees (including finances, memberships, and nominations and elections), Industry advisory board, and initiative (including PSI)
- Funded by registration fees, individual, depends on country

This is now!

COMBINE forum 2012

- Organizer: Gary Bader
- http://co.mbine.org/events/COMBINE_2012
- Twitter: #combine2012