ECCOREV – 25 septembre 2014

How to model it?

Une brève introduction à la modélisation et à la modélisation individu-centrée

Guillaume Beslon

INSA-Lyon, LIRIS UMR CNRS 5205

INRIA Grenoble – Rhône-Alpes

http://team.inria.fr/Beagle

BEAGLE

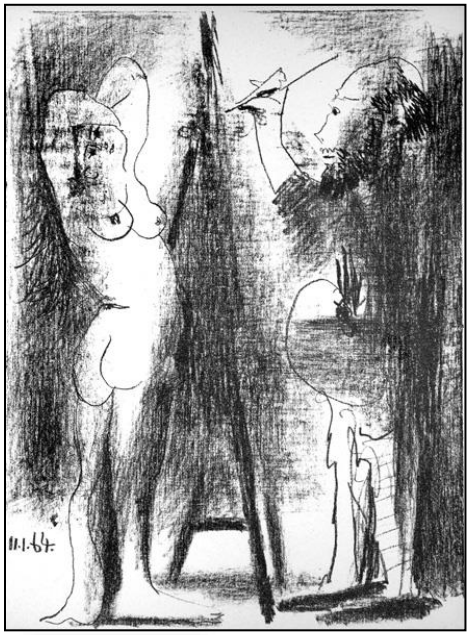
Artificial Evolution and Computational Biology

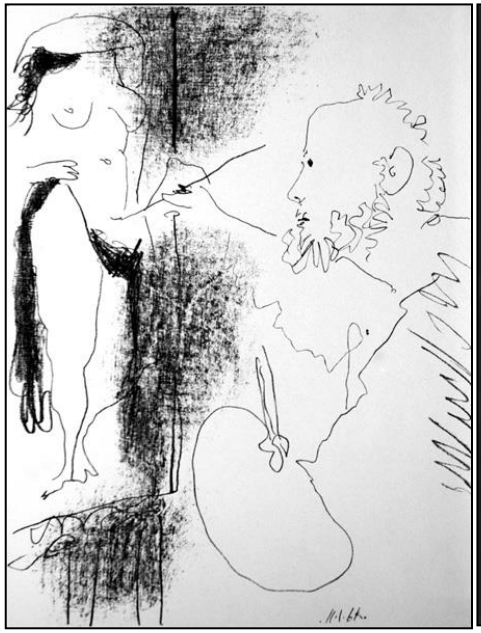
Préambule

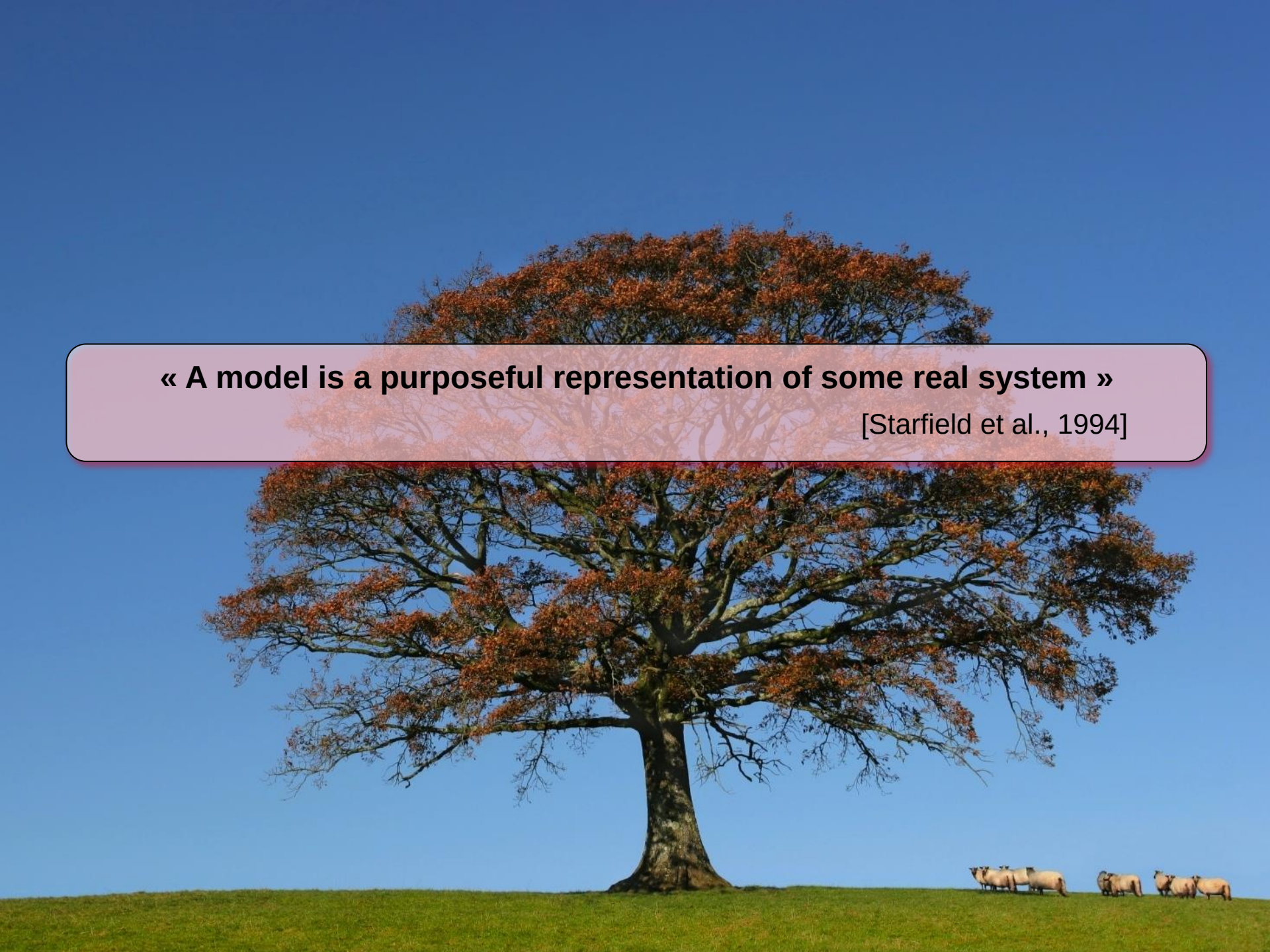
- Qui suis-je ?
  - Guillaume Beslon, guillaume.beslon@inria.fr
  - Point de vue nécessairement subjectif lié à ma pratique
  - Responsable Equipe Beagle (articifial evolution and computational biology)
  - Une assez bonne expérience des bonnes (et des mauvaises) pratiques en modélisation IC, essentiellement en biologie cellulaire et biologie évolutive
- Plan :
  1. Qu'est-ce qu'un modèle ?
  2. Qu'est-ce que la modélisation individu-centrée (« basée agents »)
  3. Un exemple de modèle individu-centré en évolution moléculaire
  4. Conclusions
- Si vous ne voulez pas écouter plus longtemps, un seul message …
  - En modélisation la technique est importante mais néanmoins secondaire par rapport à la construction du discours …

Qu’est-ce qu’un modèle ?

- Pas d’acception claire (même en science)
- Le modèle est intimement lié à l’imitation, à l’analogie, à la ressemblance
  - Mais il peut représenter aussi bien l’objet à imiter que l’imitation de l’objet ou un intermédiaire entre l’objet et l’imitation …
  - Chaque domaine/discipline à sa propre conception des modèles







« A model is a purposeful representation of some real system »

[Starfield et al., 1994]

Qu'est-ce qu'un modèle scientifique ?

« To an observer B, an object A* is a model of an object A to the extent that B can use A* to answer questions that interest him about A. »

(Marvin Minsky, 1995)

- Un modèle n'a de sens (n'est « utile ») que dans le cadre de la relation <Objet, Question, Interprète>
  - Le modèle n'est jamais une « simple » représentation d'un objet !
- Quelles sont les implications méthodologiques pour le modélisateur ?
  - Est-ce que « tout est bon ? » (Paul Feyerabend)
  - La modélisation doit être comprise dans une logique d'instrumentation
- Le modèle est donc un « instrument scientifique »
  - Il doit être utilisé comme tel
  - Mais est-il un instrument comme les autres ?

Qu'est-ce qu'un modèle scientifique ?

« To an observer B, an object A* is a model of an object A to the extent that B can use A* to answer questions that interest him about A. »

(Marvin Minsky, 1995)

- Définition très permissive : est-ce que tout est (pourrait être) modèle ?
  - Non : le modèle doit servir à répondre *scientifiquement* à une question …
  - La modélisation doit être comprise dans un contexte épistémologique …
- Le modèle est un instrument scientifique et doit être utilisé comme tel …
  - Un instrument est un objet technologique permettant de produire des faits
  - Il est accepté par une communauté scientifique

    → consensus sur l'usage de l'instrument …
  - Le modèle doit donc être considéré comme un instrument valide …
  - Pour cela il doit se conformer aux pratiques scientifiques correspondant au champ d'étude (lequel ? A ? B ? A* ?)
- Mais chaque modèle est un instrument *différent* propre à un *observateur*

Le modèle est une pièce à deux faces

- Le modèle est un instrument personnel
  - En ce sens son usage est TRES permissif …
- Le modèle est un instrument collectif
  - En ce sens son usage est TRES restrictif …
- Dans les deux cas son usage est très dangereux
  - En tant qu'instrument systématiquement nouveau, il doit être faire systématiquement ses preuves (et non faire preuve) …
  - Risque lié à l'usage personnel (preuve insuffisante ou fausse)
  - Risque lié à l'usage collectif (preuve non reconnue par la communauté)
- Or, la modélisation a toujours un caractère interdisciplinaire
  - L'usage individuel et l'usage collectif sont souvent conduits au sein de disciplines différentes …
  - Un (bon) modèle ne peut pas être construit (totalement) hors du champ disciplinaire de sa question

Un instrument personnel

« Ce qui est simple est faux, ce qui ne l'est pas est inutilisable »
(Paul Valéry)

- Par définition le modèle est une simplification. Il ne fait *jamais* preuve
  - Mais ça n'interdit pas son utilité (au contraire !)
  - *The decisive thing with modelling is not the model per se, but what the model and working with the model does to our mind* [V. Grimm, 1999]
  - Critères de qualité d'un modèle : l'« utilité » [J.-M. Legay, 1973] ou la « pertinence » [J.-L. Le Moigne, 1977]
  - Le modèle n'est donc pas un résultat, un objectif scientifique *en soi* ; il est indissociable de sa construction et de son utilisation
- C'est le modélisateur qui incarne le lien entre le modèle et son objet
  - Le modélisateur doit posséder les compétences nécessaires pour interpréter son modèle à toutes les étapes (construction, utilisation, communication)
  - « Le hasard ne sourit qu'aux esprits préparés » (L. Pasteur)

Un instrument collectif

« Il y a peu de controverses *entre simulateurs* car il y a peu de travail collectif. Les simulateurs sont rassemblés par l'équipement informatique qui leur est nécessaire, mais ils fonctionnent plutôt à la manière de petits artisans : chacun son problème, son modèle, son programme »

[Stengers & Bousaude-Vincent, 2003]

- Le modèle est un instrument personnel mais qui doit autoriser les échanges avec le collectif …
  - Prouver c'est convaincre …
  - Convaincre qui ? De quoi ?
- Chaque champ d'application va exiger du modèle (et du modélisateur) qu'il se plie aux règles (implicites) du domaine
  - Sous peine d'être immédiatement déconsidéré
  - Un modèle peut-il être un instrument valide puisqu'il est toujours un instrument ad-hoc (construit pour répondre à une question) ?
  - Pour convaincre, le modèle doit d'abord conforter …
- Le passage au collectif change le statut du modèle

Comment « classer » les modèles ?

- Les typologies classiques sont basées sur le type de représentation
  - Modèles graphiques, mathématiques, statistiques, numériques, informatiques, formels, computationnels, individu-centré, basé-agents, …
- Mais si un modèle est une représentation *utile* de quelque-chose …
  - → Utile à quoi ? Utile à qui ? Utile pourquoi ? Utile quand ?
  - → Une typologie des questions est plus appropriée
- En science, un modèle peut être utile de plusieurs manières différentes
  - Prédire, expliquer, **fabriquer**, transmettre, partager, distinguer, …
    - → Modèle prédictif (prédire à partir de la théorie)
    - → **Modèle explicatif (construire la théorie à partir de l'observation)**
    - → Modèle normatif (spécifier un « objet » avant sa réalisation)
    - → Modèle pédagogique (idéaliser un objet pour communiquer)
- Chaque type de modèle/question pose des problèmes spécifiques …

Qu'est-ce qu'un modèle (explicatif) individu-centré ?

- Les modèles computationnels explorent les propriétés d'un système par le calcul (formel ou numérique)
  →Les modèles individu-centrés sont des modèles computationnels basés sur une *description explicite* des entités qui composent un système et sur l'*observation* des propriétés « émergentes » du système
- Modélisation « ascendante » ("Bottom-Up") :
  - Description du système au niveau local au moyen d'« un » formalisme
  - Simulation computationnelle du système
  - Observation et analyse des résultats (aux deux échelles)

"*In agent-based modeling (ABM), a system is modeled as a collection of autonomous decision-making entities called agents. Each agent individually assesses its situation and makes decisions on the basis of a set of rules. […] Repetitive competitive interactions between agents are a feature of agent-based modeling, which relies on the power of computers to explore dynamics out of the reach of pure mathematical methods.*"

[Bonabeau, 2002]

Modèles Basés Agents (ABM)

Field: Research Areas	Record Count	% of 9867	Bar Chart
COMPUTER SCIENCE	5200	52.701 %	
ENGINEERING	2461	24.942 %	
BUSINESS ECONOMICS	840	8.513 %	
OPERATIONS RESEARCH MANAGEMENT SCIENCE	748	7.581 %	
MATHEMATICS	698	7.074 %	
ENVIRONMENTAL SCIENCES ECOLOGY	656	6.648 %	
AUTOMATION CONTROL SYSTEMS	509	5.159 %	
SOCIAL SCIENCES OTHER TOPICS	439	4.449 %	
TELECOMMUNICATIONS	393	3.983 %	
PHYSICS	383	3.882 %	

Modèles Individu-Centrés (IBM)

Field: Research Areas	Record Count	% of 4056	Bar Chart
ENVIRONMENTAL SCIENCES ECOLOGY	1799	44.354 %	
MARINE FRESHWATER BIOLOGY	379	9.344 %	
EVOLUTIONARY BIOLOGY	373	9.196 %	
LIFE SCIENCES BIOMEDICINE OTHER TOPICS	348	8.580 %	
COMPUTER SCIENCE	292	7.199 %	
FISHERIES	283	6.977 %	
MATHEMATICAL COMPUTATIONAL BIOLOGY	282	6.953 %	
OCEANOGRAPHY	240	5.917 %	
GENETICS HEREDITY	225	5.547 %	
SCIENCE TECHNOLOGY OTHER TOPICS	200	4.931 %	

→ ABM & IBM … Même approche mais issus de domaines scientifiques différents

→ Attention : ce n'est pas sans conséquences …

Systèmes Multi-Agents (MAS)

Field: Research Areas	Record Count	% of 16664	Bar Chart
COMPUTER SCIENCE	11760	70.571 %	
ENGINEERING	5820	34.926 %	
AUTOMATION CONTROL SYSTEMS	3206	19.239 %	
OPERATIONS RESEARCH MANAGEMENT SCIENCE	1032	6.193 %	
TELECOMMUNICATIONS	909	5.455 %	
ROBOTICS	899	5.395 %	
MATHEMATICS	683	4.099 %	
BUSINESS ECONOMICS	517	3.102 %	
ENERGY FUELS	306	1.836 %	
INSTRUMENTS INSTRUMENTATION	291	1.746 %	

→ Les SMA sont des modèles de calcul (modèles normatifs)

→ La proximité thématique entre SMA et ABM entraîne de nombreuses confusions (il n'y a aucune raison d'utiliser un SMA comme formalisme pour implémenter un ABM)

Qu'est-ce qu'un agent/individu ?

- Trois définitions différentes :
  - Technique : un élément logiciel persistant et autonome remplissant une tache particulière
  - Anthropomorphique : une entité capable de percevoir son environnement et d'agir rationnellement en fonction d'un but
  - Méthodologique : l'unité conceptuelle d'intérêt, spécifiée par le modélisateur, définissant la frontière entre ce qui défini et ce qui est observé
- L'agent est « seulement » l'unité de description au niveau microscopique
  - Il n'a pas nécessairement de réalité technique …
- L'agentification est un point de vue sur le système ; c'est une prise de position du modélisateur
  - Déterminée avant tout par la question scientifique et le domaine d'étude …
  - L'aspect technologique est « secondaire » !

"*Agent-Based Model is a mindset more than a technology.*"

[Bonabeau, 2002]

Construction un modèle individu-centré

- Une fois les agents identifiés, la construction d'un modèle individu-centré est techniquement assez facile …

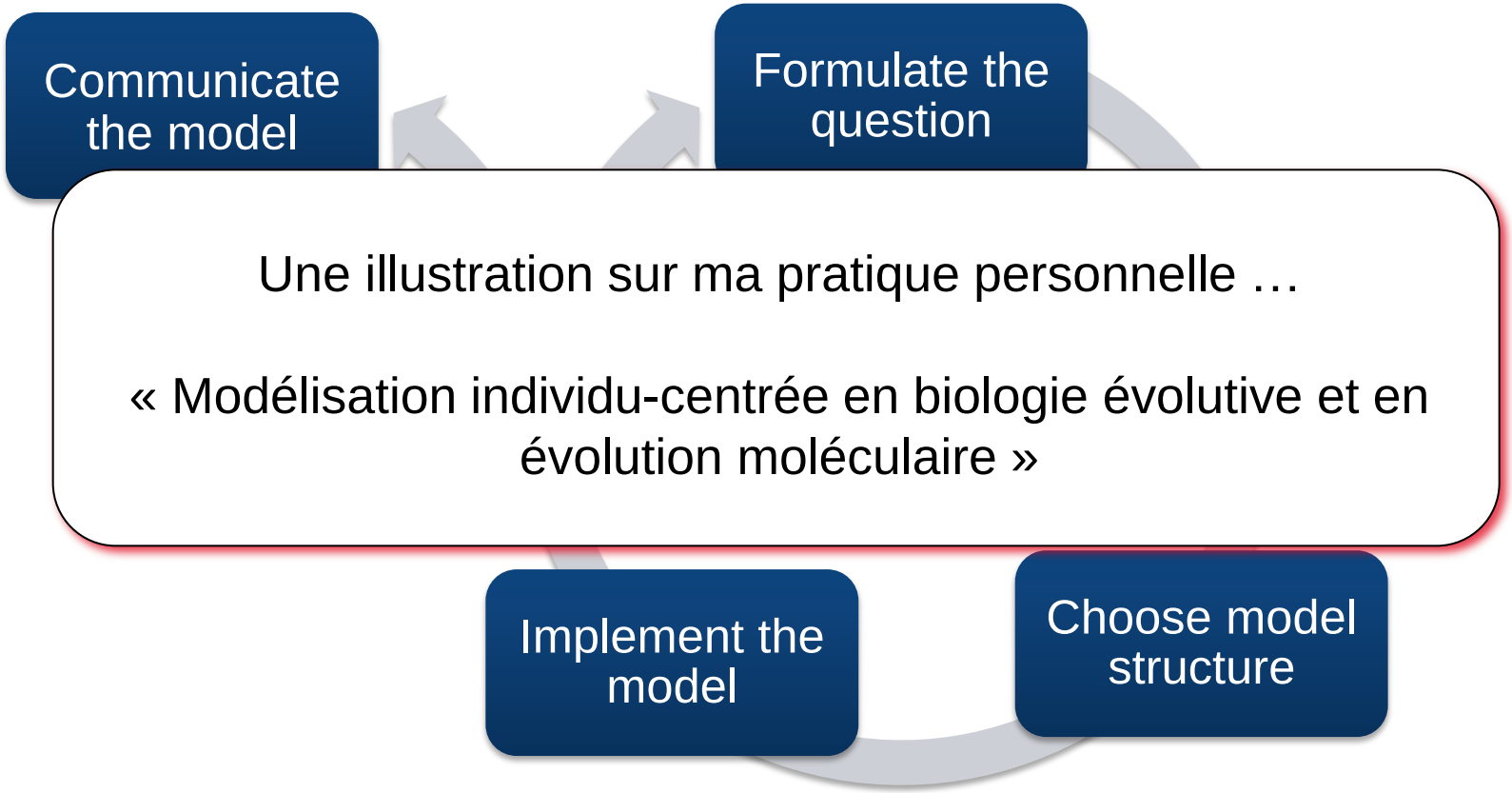
  - Très nombreux tutoriaux sur différents langages/plateformes/…
  - Souvent inspirés de l'ingénierie logiciel ou des SMA (cf. [North & Macal, 2006])
- Certains points peuvent s'avérer sensibles (the devil is in the details)

  → Modélisation du temps, de l'espace, de l'aléa, …
- Mais l'approche technologique cache la vraie subtilité …

Dans la vie scientifique, les problèmes ne se posent pas d'eux-mêmes. C'est précisément ce sens du problème qui donne la marque du véritable esprit scientifique. Pour un esprit scientifique, toute connaissance est une réponse à une question. S'il n'y a pas eu de question, il ne peut y avoir de connaissance scientifique. Rien ne va de soi. Rien n'est donné. Tout est construit.

[Gaston Bachelard, 1947]

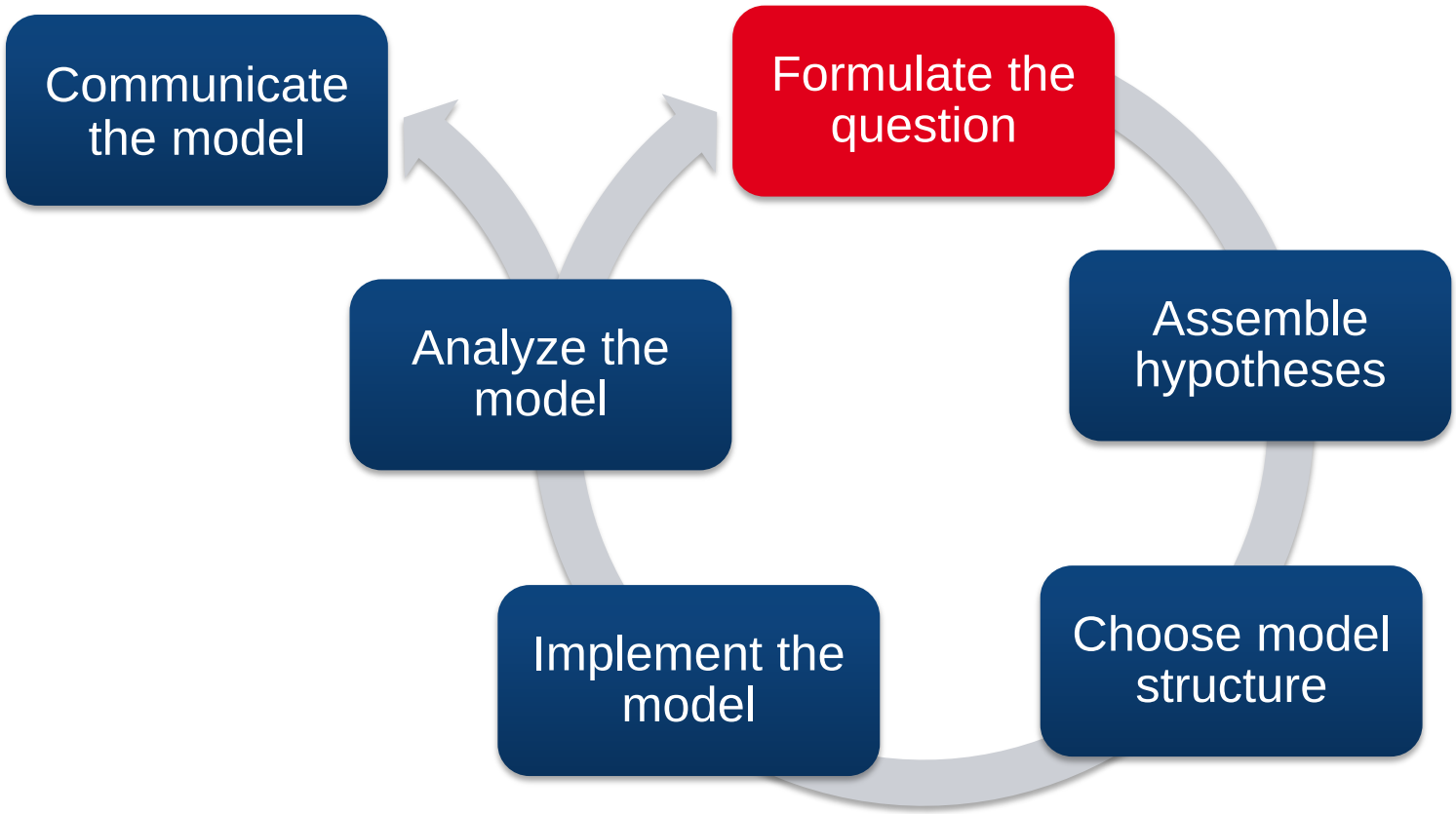
Le cycle de la modélisation individu-centrée



Une illustration sur ma pratique personnelle …

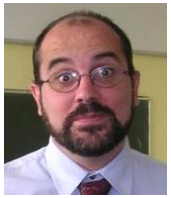
« Modélisation individu-centrée en biologie évolutive et en évolution moléculaire »

From [Railsback & Grimm, 2012]



From [Railsback & Grimm, 2012]

Quelle est ma question ?

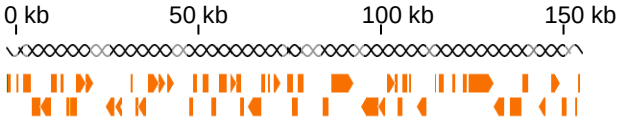


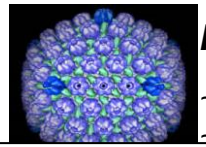
Homo sapiens

~3 billions bp
~25 000 genes

Neisseria meningitidis

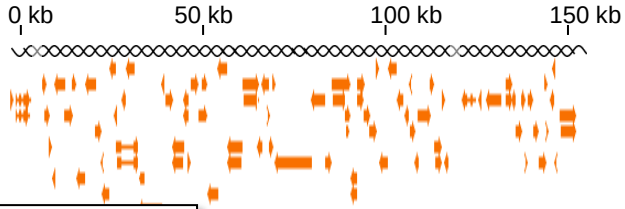
~2 millions bp
~2 000 genes



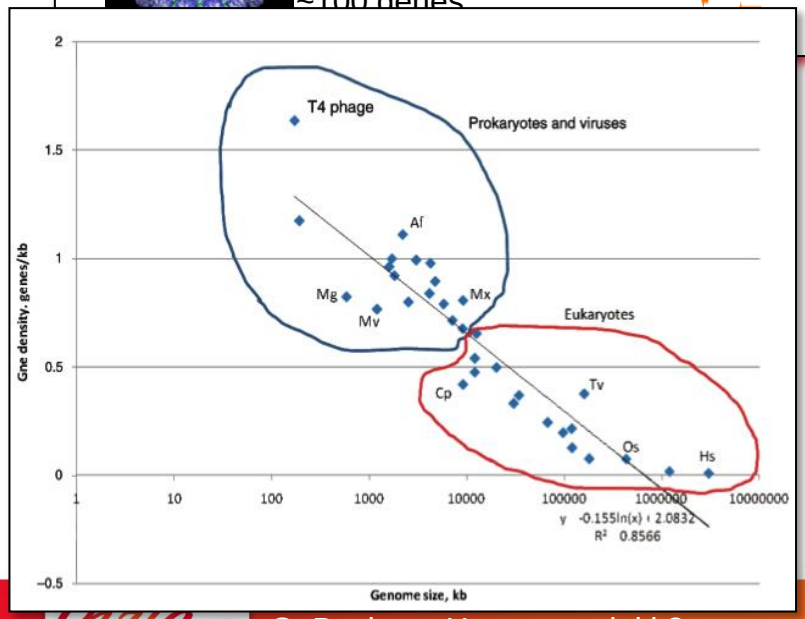


Herpes HSV-1

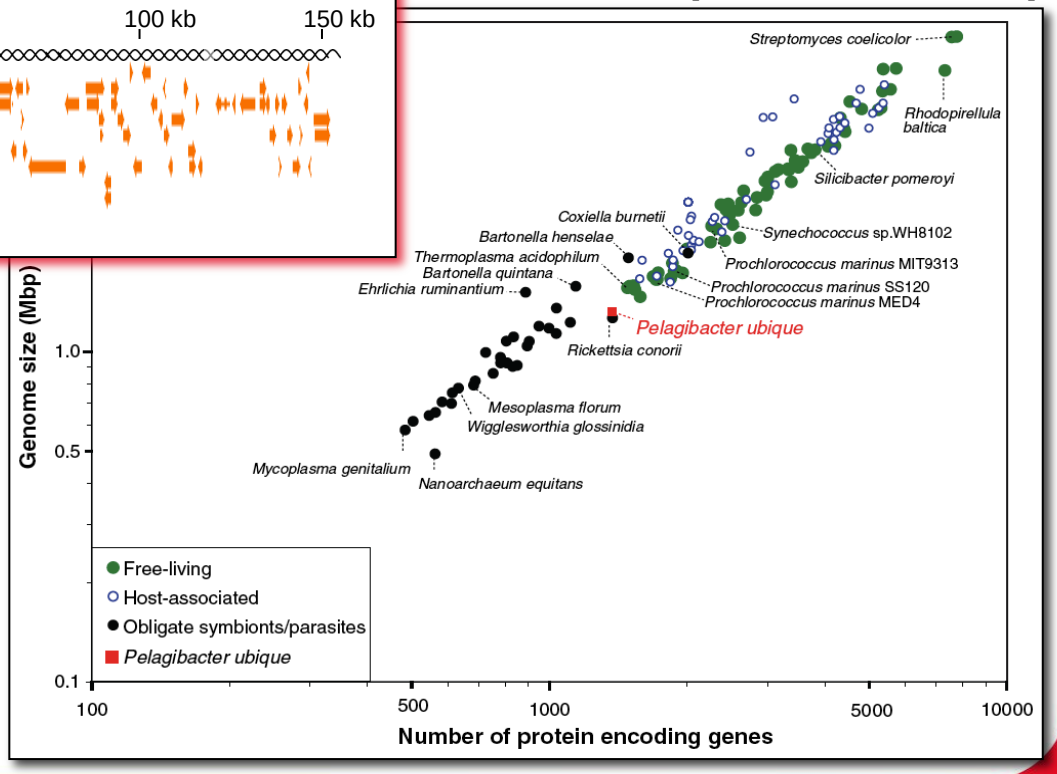
~150 000 bp
~100 genes

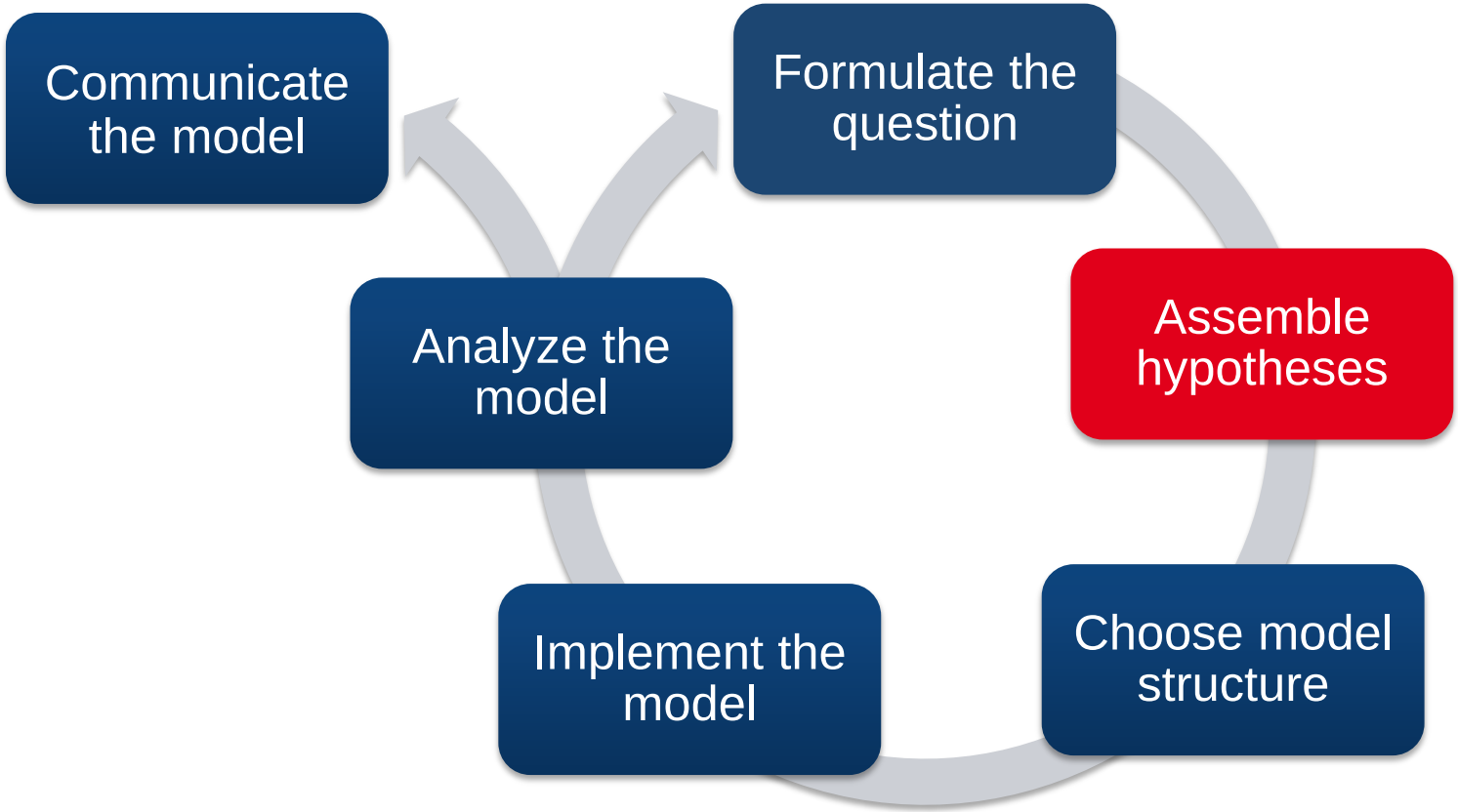


[Giovanonni et al., 2005]



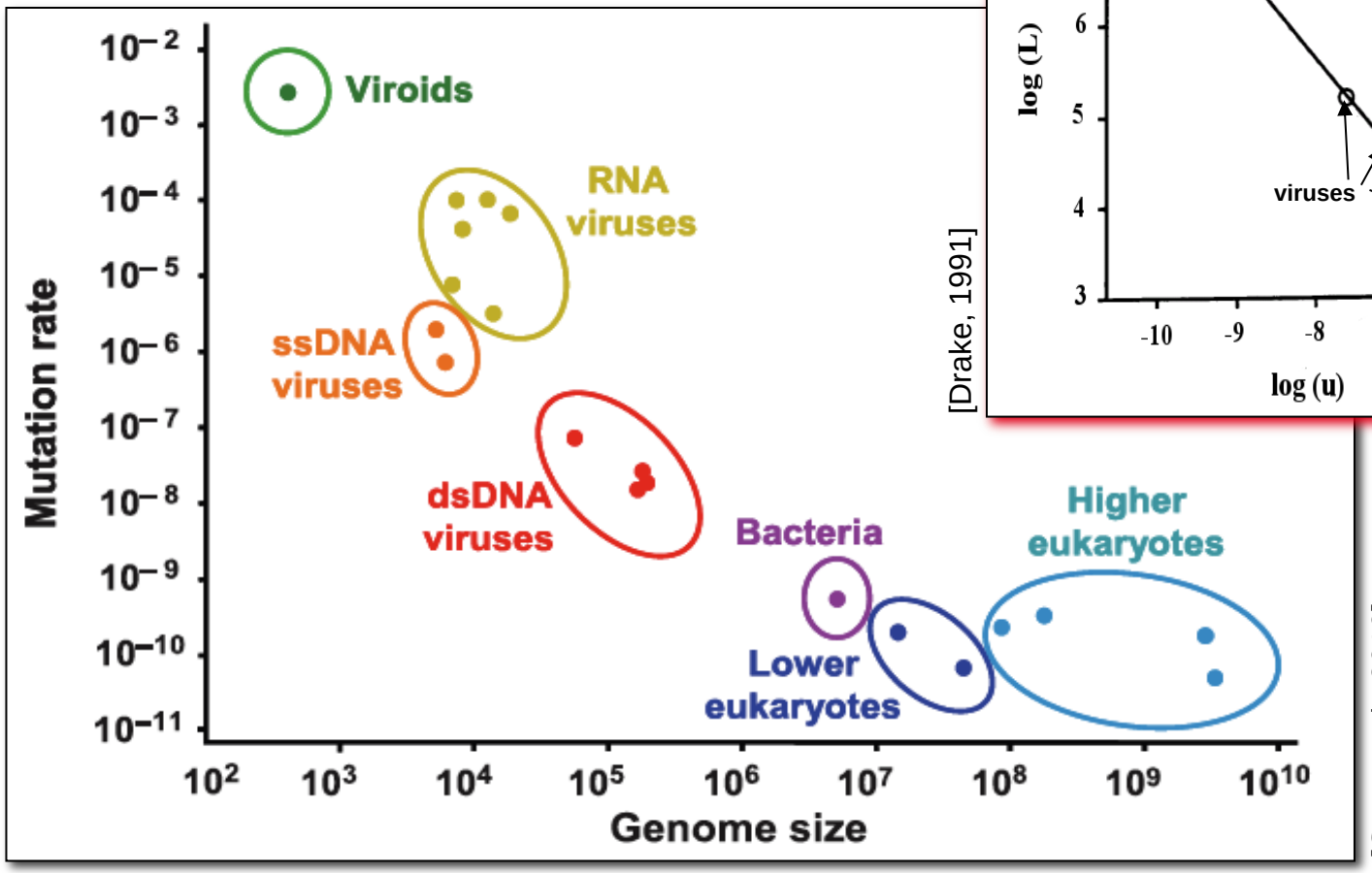
[Koonin, 2009]



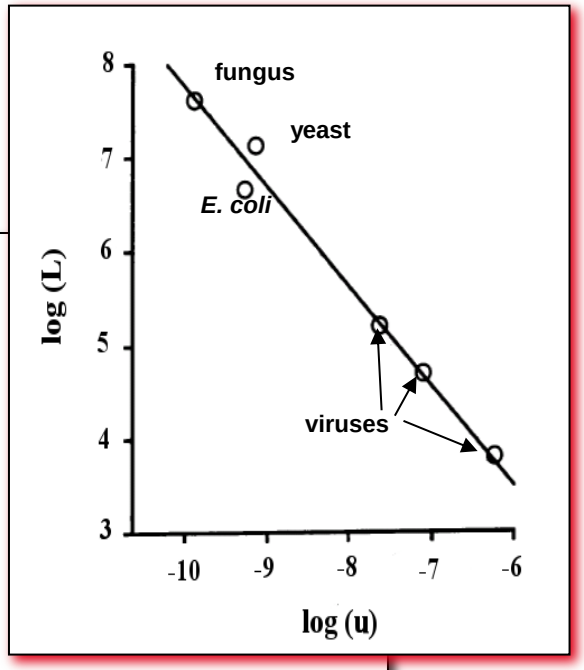


From [Railsback & Grimm, 2012]

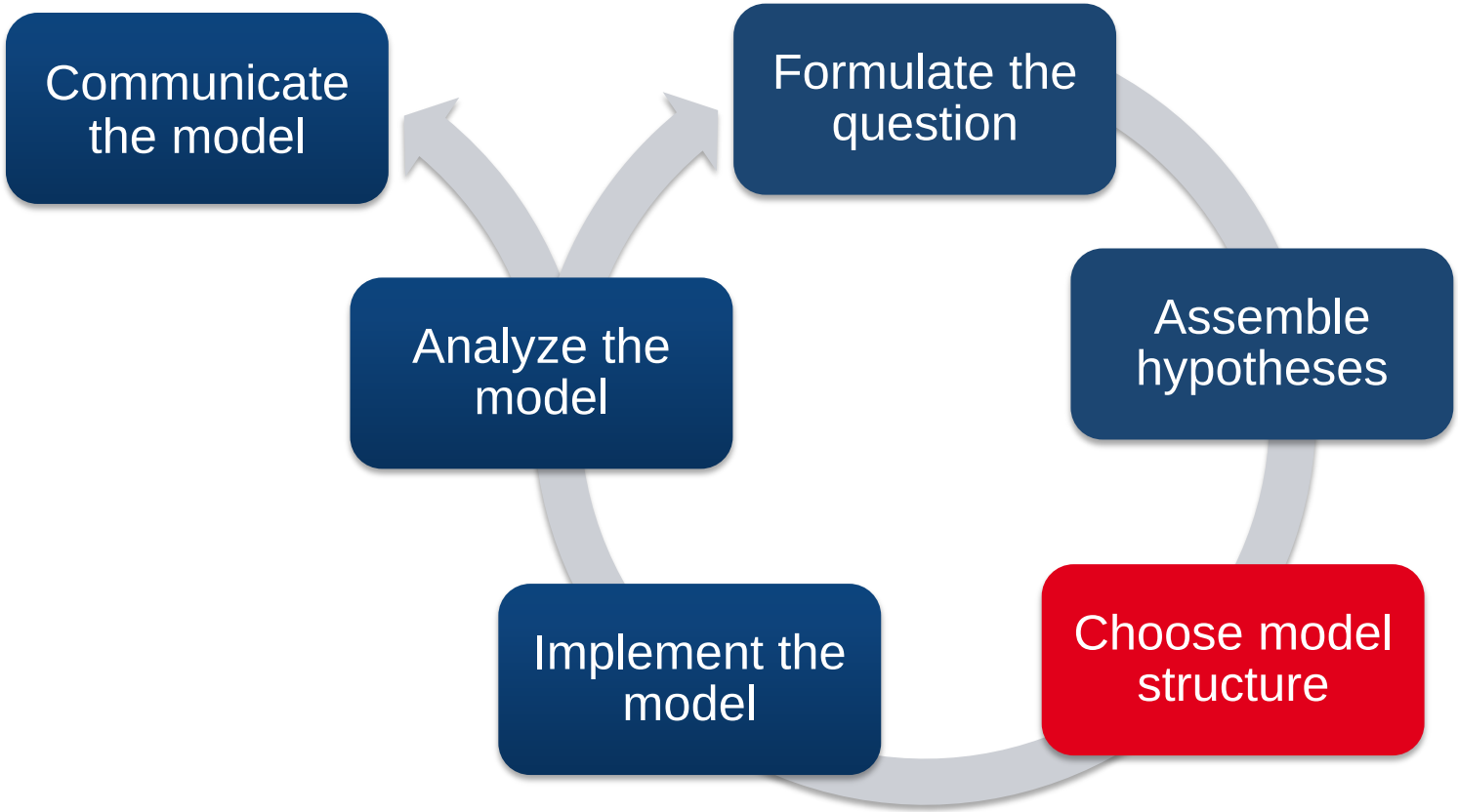
Quelle(s) hypothèse(s) ?



[Gago et al., 2010]

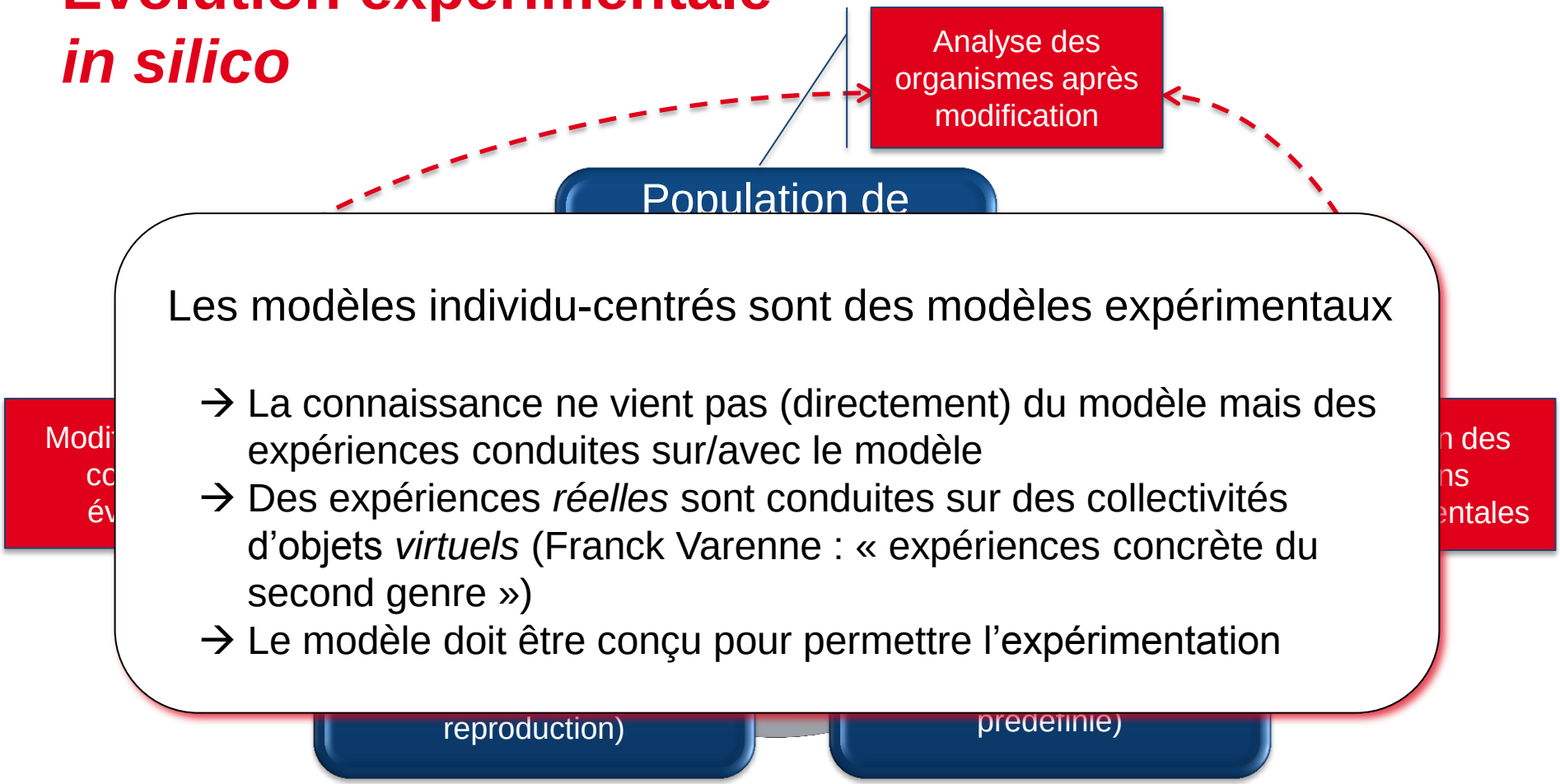


[Drake, 1991]



From [Railsback & Grimm, 2012]

Evolution expérimentale *in silico*



Les modèles individu-centrés sont des modèles expérimentaux

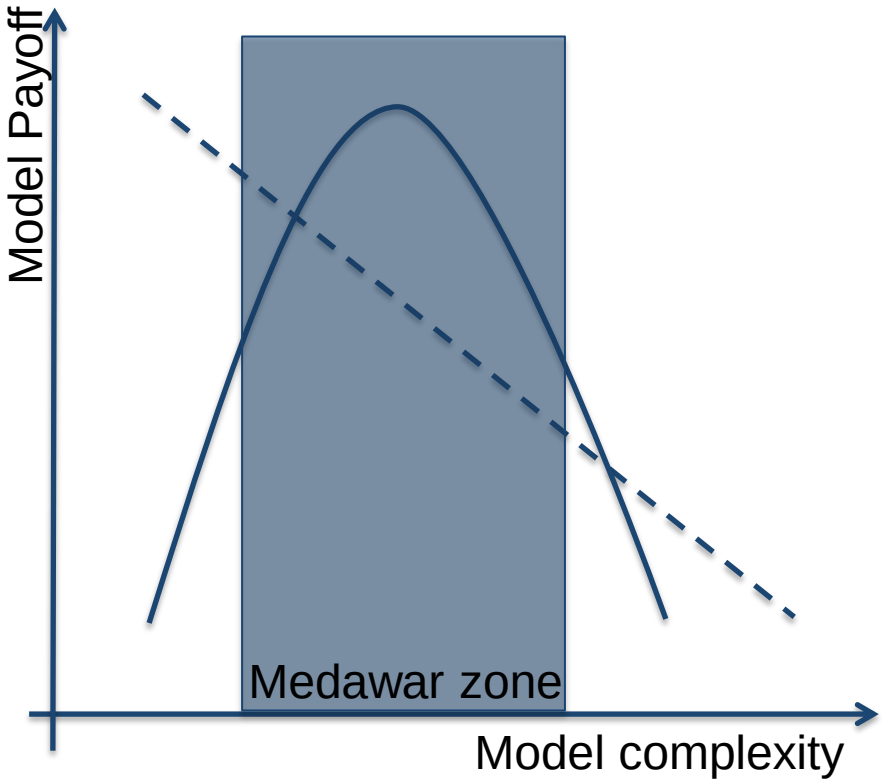
→ La connaissance ne vient pas (directement) du modèle mais des expériences conduites sur/avec le modèle
→ Des expériences *réelles* sont conduites sur des collectivités d'objets *virtuels* (Franck Varenne : « expériences concrète du second genre »)
→ Le modèle doit être conçu pour permettre l'expérimentation

"*Digital organisms make it easy to conduct repeatable, controlled experiments, which have a perfect genetic fossil record*"

[Adami, *Nat. Rev. Genet.*, 2006].

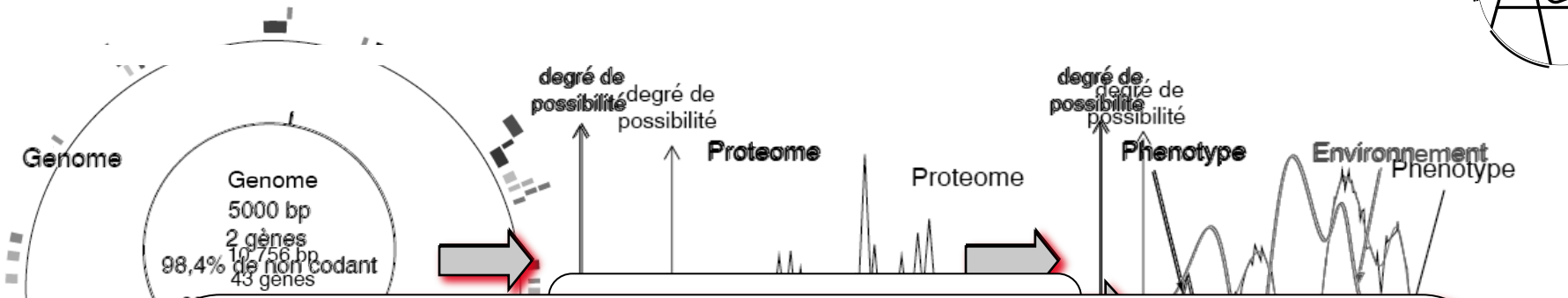
Quelle complexité pour le modèle/les agents ?

- Modèles analytiques
  - Il est difficile d'accroitre la complexité du modèle
- Modèles individu-centrés
  - Les possibilités du modèle semblent (souvent) illimitées
  - Il est difficile de limiter la complexité du modèle !
- La complexité des agents dépend de la question, des expériences, des opportunités de réutilisation et, surtout, de l'expressivité vis à vis de la communauté
  - La complexité du modèle doit être guidée par la construction du discours, pas par une volonté d'épurer a priori



From: [Grimm & Railsback, 2005]

Aevol: une plate-forme d'évolution expérimentale in silico



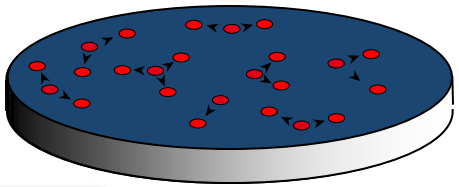
Dans Aevol, les organismes possèdent un génome structuré (nombre de gènes variable, taille variable, structures opéroniques, séquences non-codantes, …).

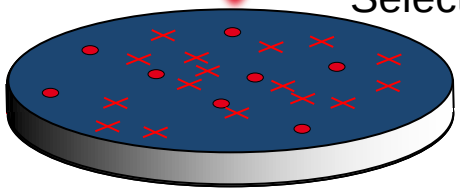
Aevol permet de tester expérimentalement la dépendance de la structure du génome aux conditions évolutives (typiquement les taux de mutation).

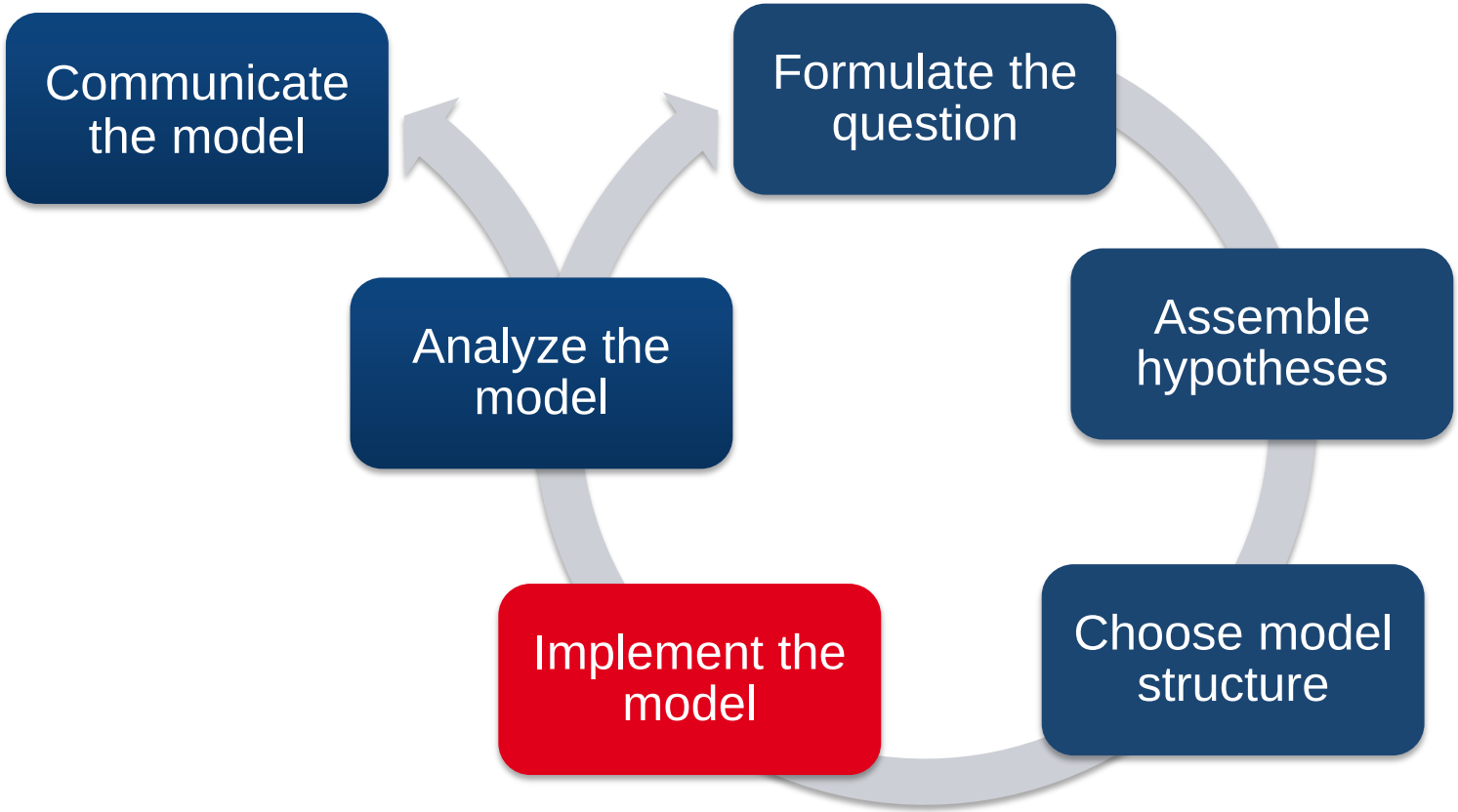
Réplication (mutations, réarrangements)

Réplication

Sélection

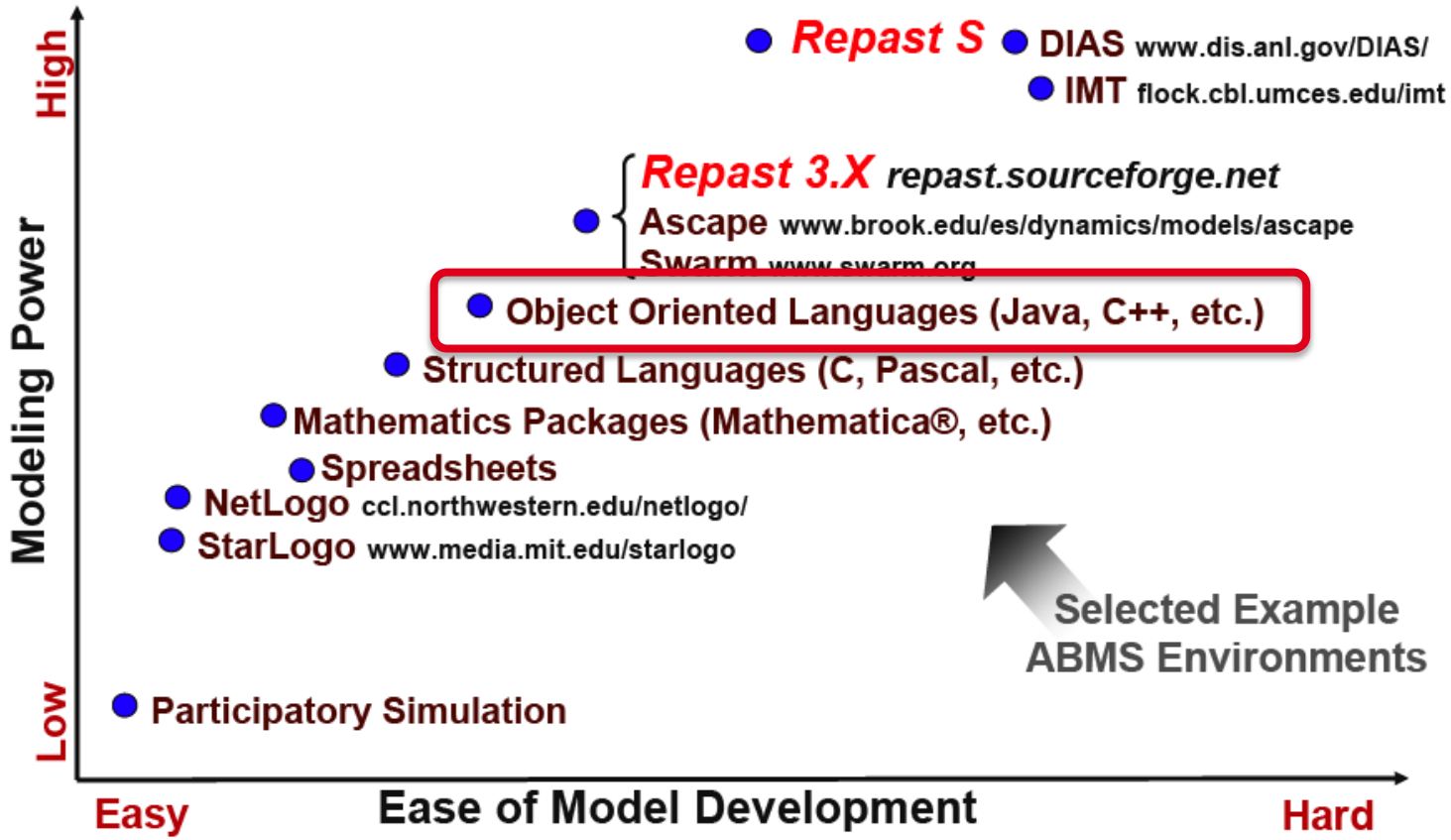




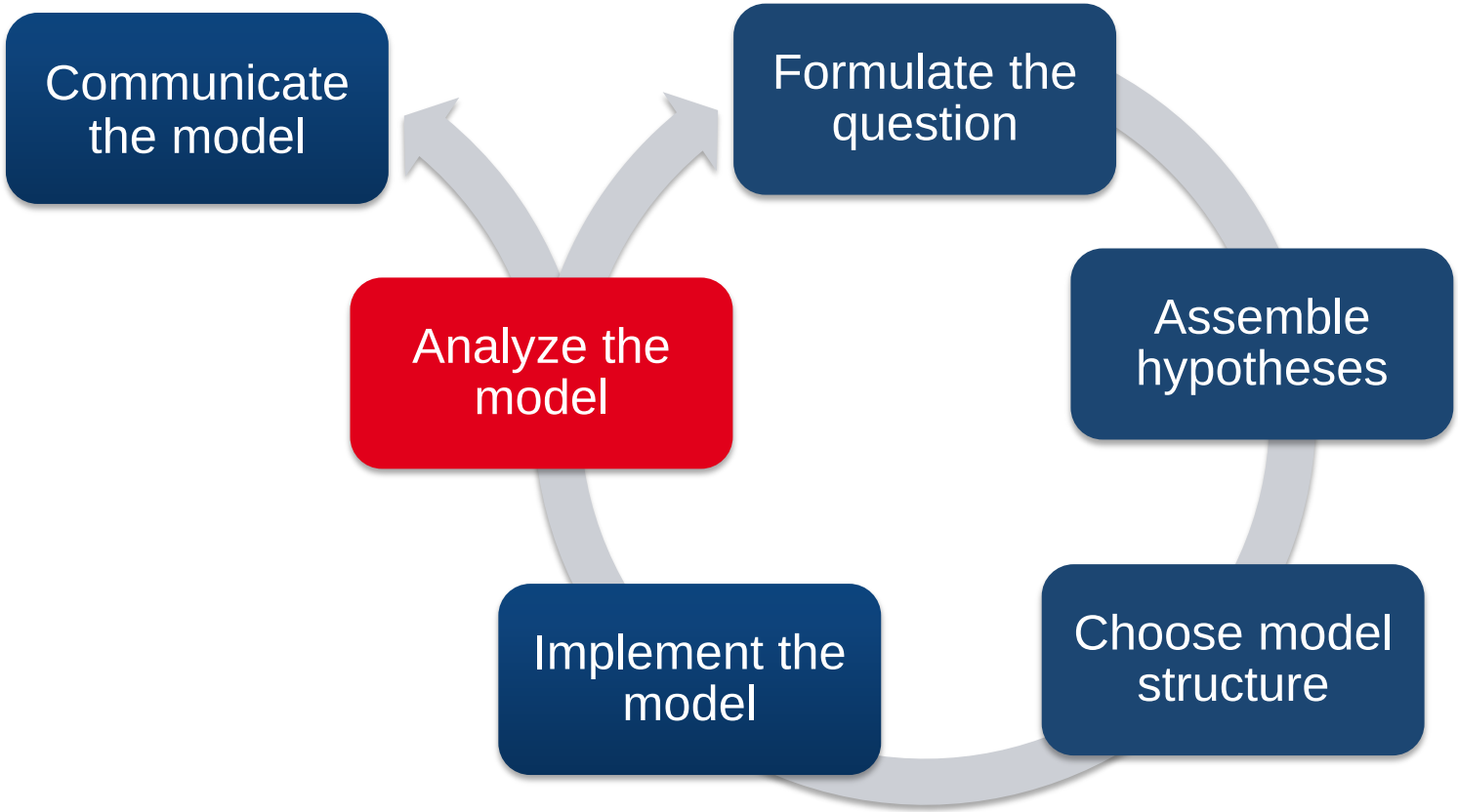


From [Railsback & Grimm, 2012]

Implémentation



[North & Macal, 2006]

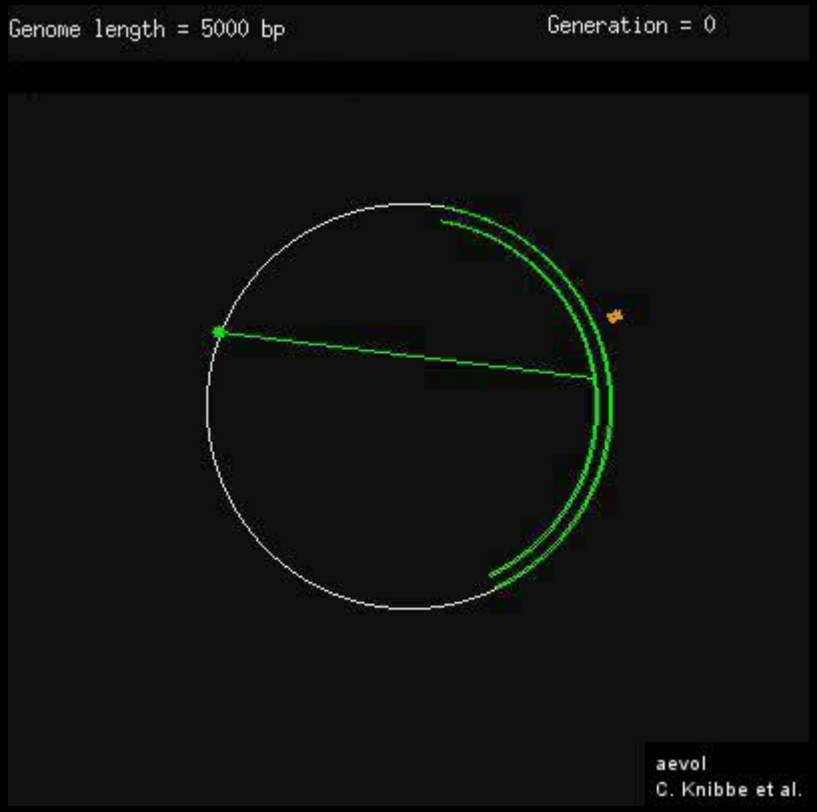


From [Railsback & Grimm, 2012]

L'analyse du modèle est un problème crucial

- Les modèles individu-centrés sont souvent complexes et très paramétrés
  - Beaucoup de paramètres sont implicites …
  - Aevol : 53 paramètres explicites !
  - Inter-dépendances ; non-linéarités ; forte dépendance aux conditions initiales ; comportements souvent partiellement stochastiques …
- Les modèles individu-centrés sont souvent lents
  - Importants besoins en ressources de calcul
  - Impossibilité pratique de tester tous les paramètres
- La construction du modèle est souvent moins couteuse que son utilisation…
  - Estimer le temps expérimental (et les ressources) dès le début du projet
  - L'implémentation du modèle doit faciliter la conduite des expériences
- Les modélisateurs doivent importer les « bonnes pratiques » des sciences expérimentales (cahier de manip, tests statistiques, …)

Aevol : observation de l’évolution en action …

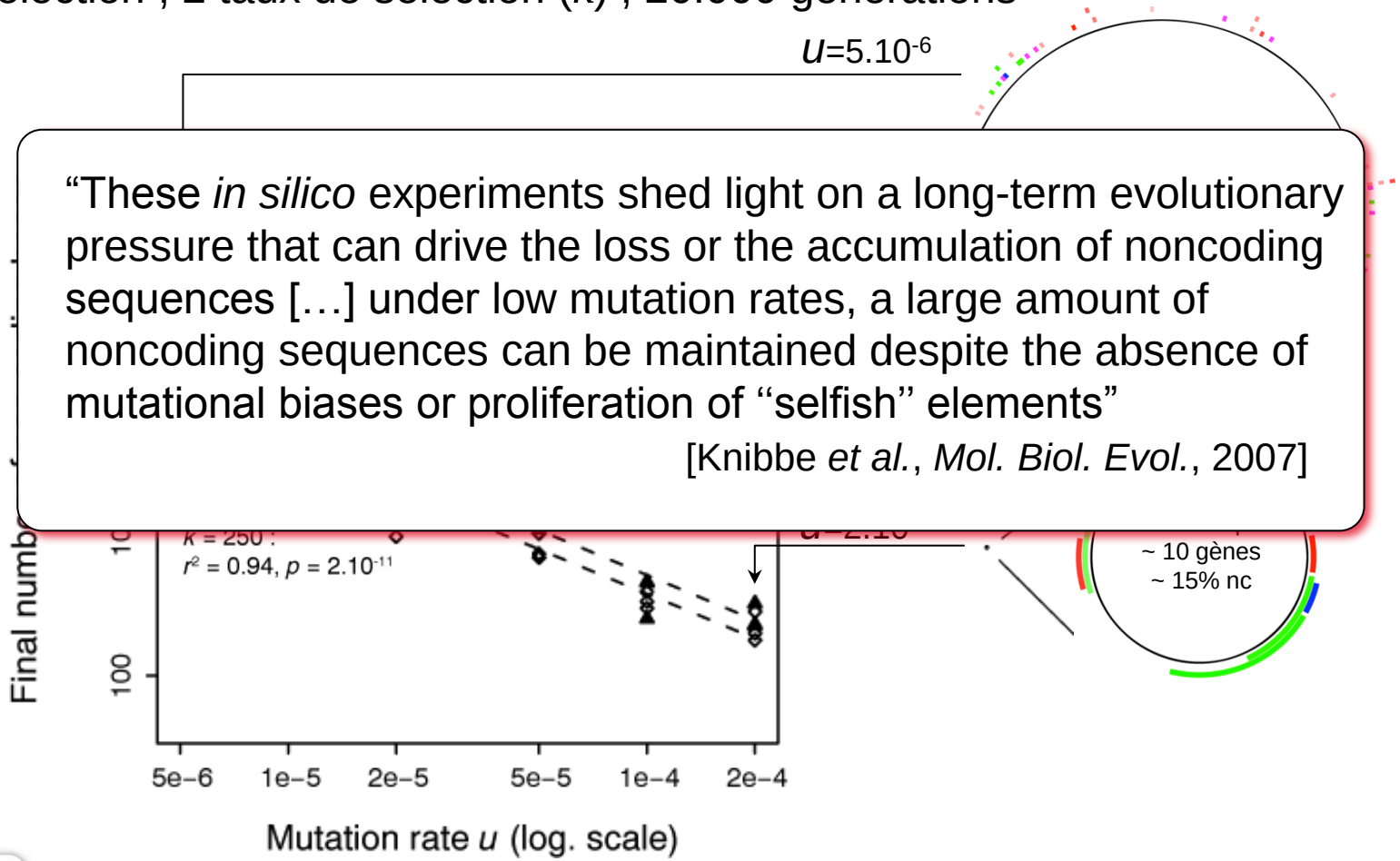


Forts taux de mutation : 2.10^{-4} mut/bp/gen

Faibles taux de mutation : 5.10^{-6} mut/pb/gen

Etude expérimentale

- Etude expérimentale de l'effet des taux de mutation et de rérrangement chromosomiques : 6 taux de mutation (u = 5.10^{-6} to 2.10^{-4} mut/bp/gen); 2 modes de sélection ; 2 taux de selection (k) ; 20.000 générations

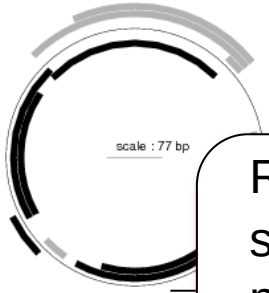


"These *in silico* experiments shed light on a long-term evolutionary pressure that can drive the loss or the accumulation of noncoding sequences […] under low mutation rates, a large amount of noncoding sequences can be maintained despite the absence of mutational biases or proliferation of "selfish" elements"

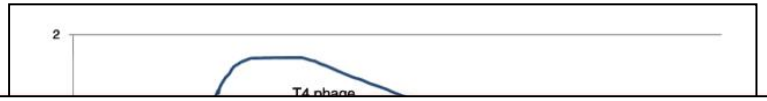
[Knibbe *et al.*, *Mol. Biol. Evol.*, 2007]

[Knibbe *et al.*, *Mol. Biol. Evol.* 2007]

Impact sur la densité du génome et sur la transcription



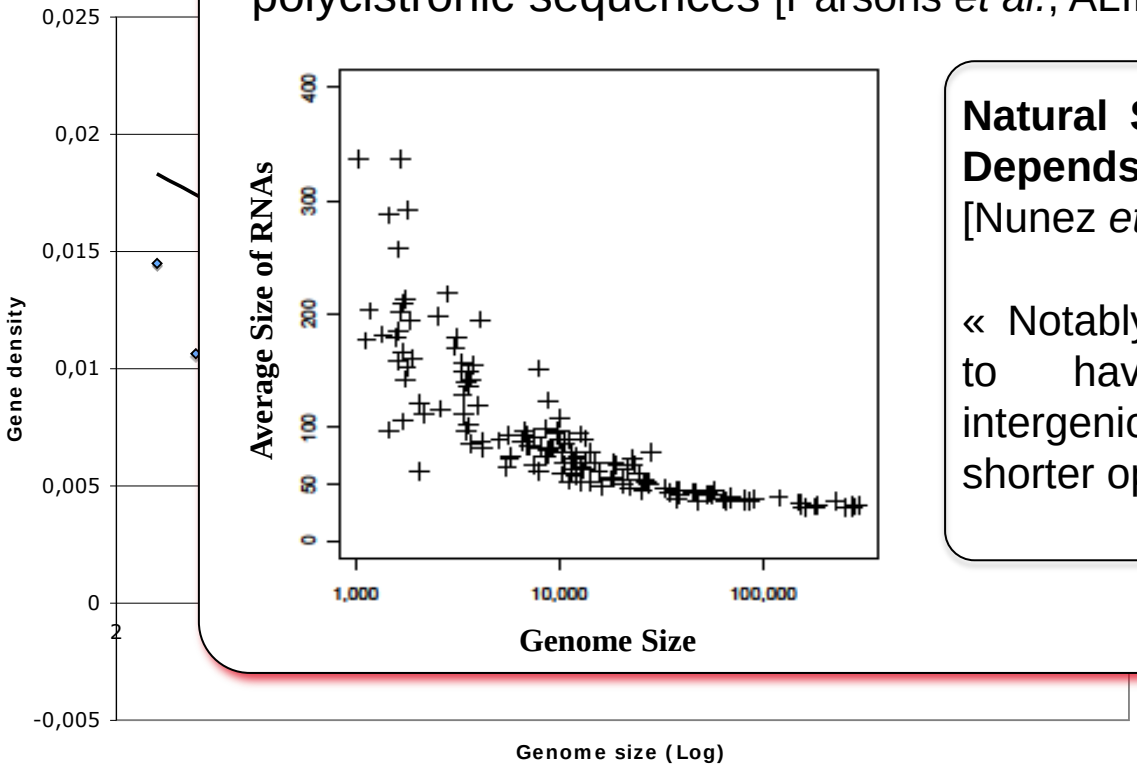
Genome size : 1,1 kbp
Excess DNA : 19%
Genes : 15



Reduction of genome size imposes an efficient packing. Hence short genomes have less ncRNAs, longer cRNAs and more polycistronic sequences [Parsons *et al.*, ALIFE XII, 2010].

Natural Selection for Operons Depends on Genome Size [Nunez *et al.*, 2013]

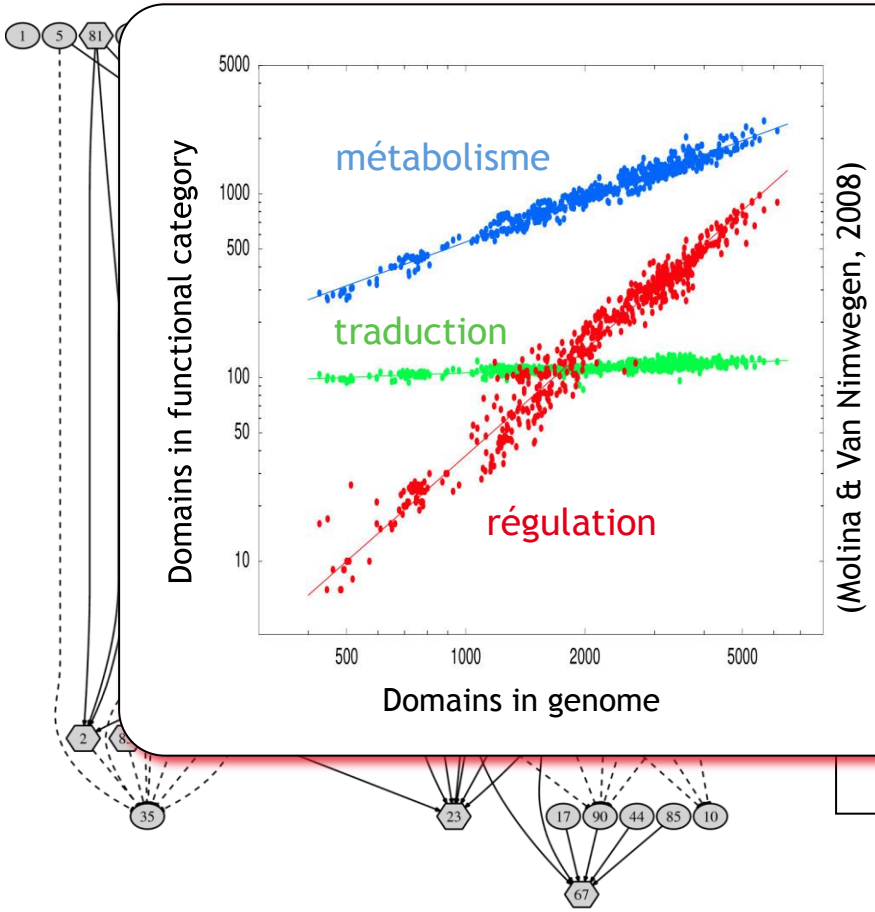
« Notably, larger genomes tend to have relatively larger intergenic regions and fewer and shorter operons. »



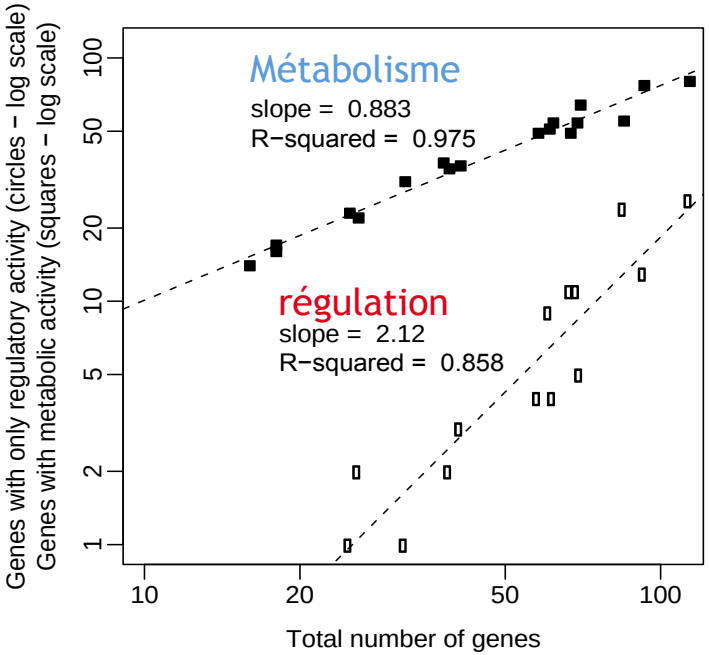
ze : 249 kbp
A : 98%
73
1236
As : 71
2
RNA size (av) : 87 bp

Impact impact sur les réseaux de régulation

$\mu = 5.10^{-6}$



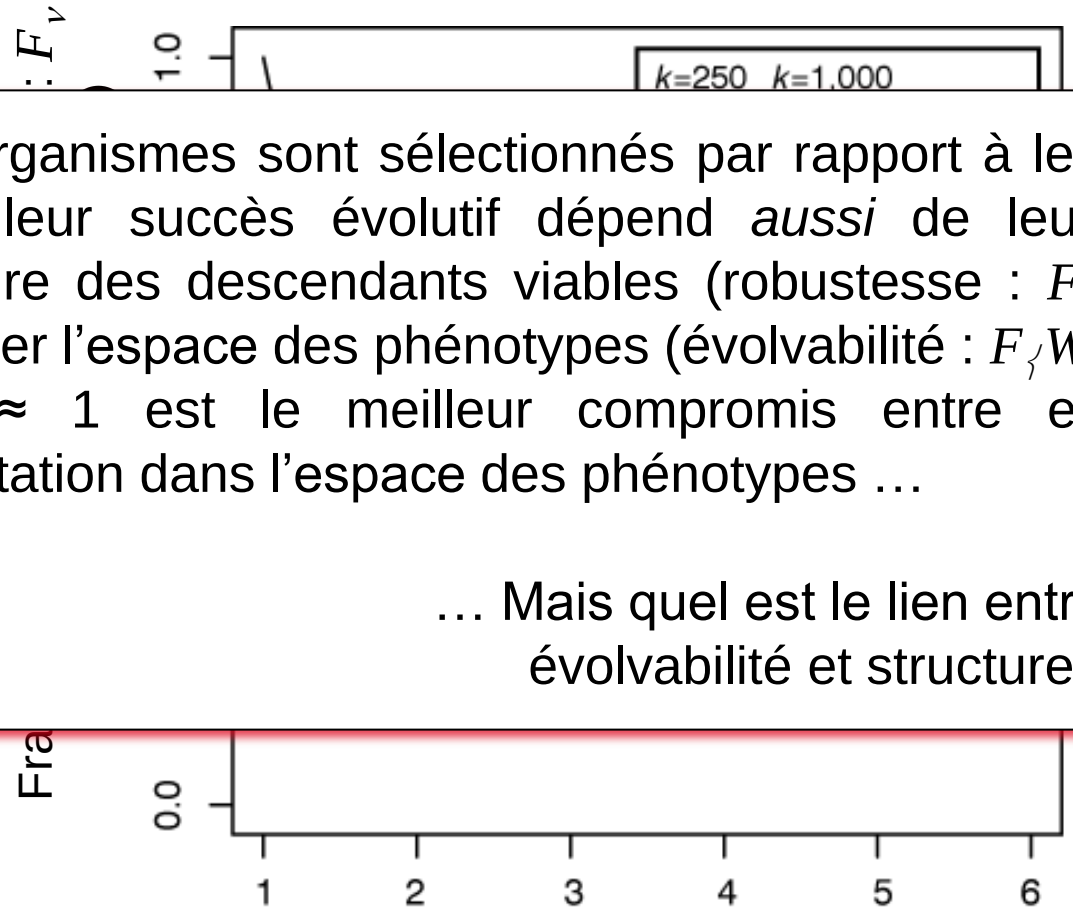
(Molina & Van Nimwegen, 2008)



[Beslon *et al.*, BioSystems 2010]

- Aevol permet de reproduire de nombreuses caractéristiques génomiques en modifiant un seul des paramètres de la simulation (le taux de mutation)
  - → Parcimonie …
  - → Aevol est-il un bon modèle ?
- « Prédire n'est pas expliquer » (R. Thom)
  - → Retour à la phase d'analyse …

Mesure de la reproduction neutre du meilleur individu à *t* = 20000

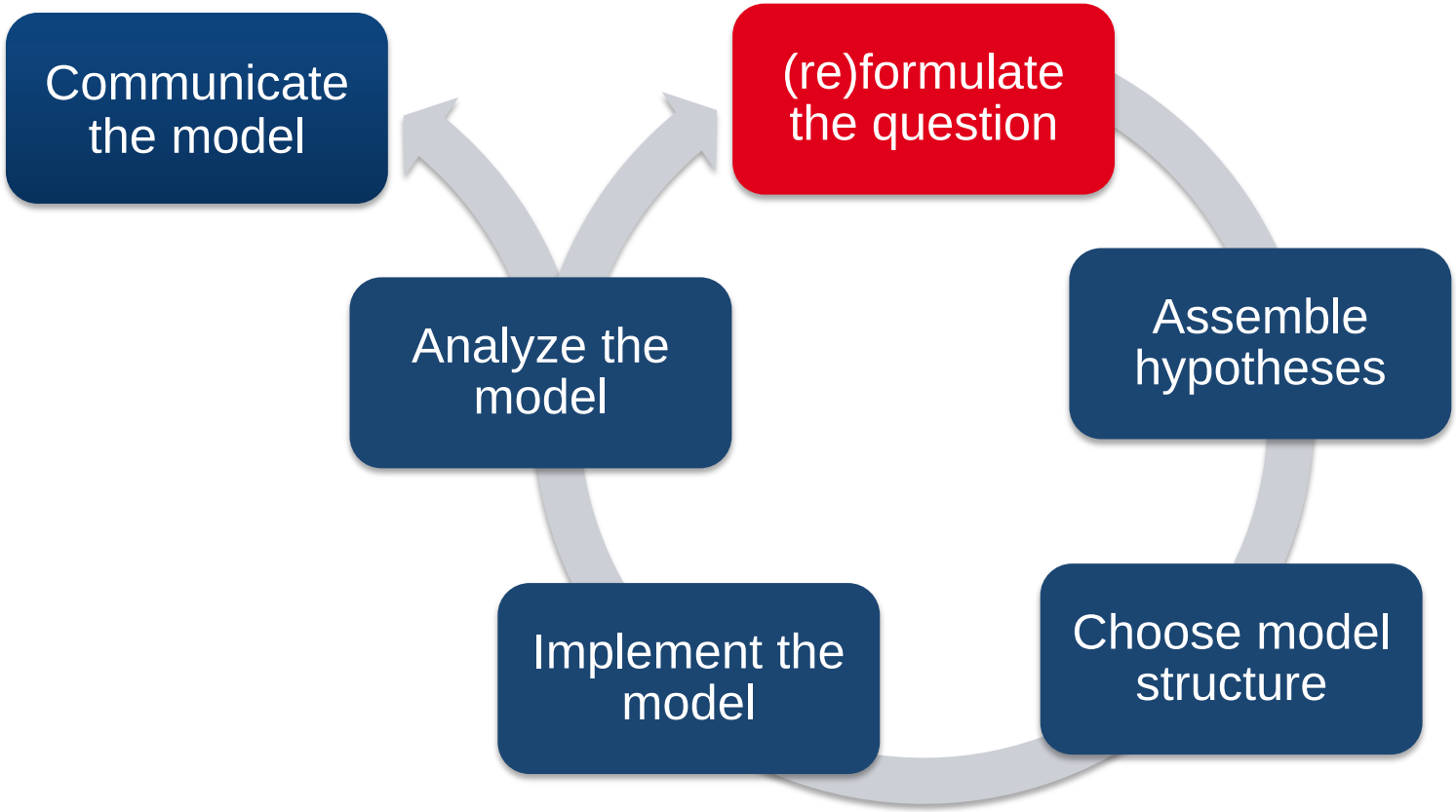


Les organismes sont sélectionnés par rapport à leur « fitness » mais leur succès évolutif dépend *aussi* de leur capacité à produire des descendants viables (robustesse : $F_{\wr}W > 1$) et à explorer l'espace des phénotypes (évolvabilité : $F_{\wr}W < 1$)…
$F_{\wr}W \approx 1$ est le meilleur compromis entre exploration et exploitation dans l'espace des phénotypes …

… Mais quel est le lien entre robustesse, évolvabilité et structure génomique ?

Nombre d'essais reproductifs : W
(dépend des rapports de fitness)

Le cycle de la modélisation individu-centrée



From [Railsback & Grimm, 2012]

Re-mathématisation du modèle …

- Constr
  - La
- $\tilde{F}_{\square}$: Pr u
  génom ue de
  chaqu
- $\tilde{\square}_i$: Pr ion e la
  structu

$$\begin{cases} \tilde{\nu}_{\text{ponct}} \\ \tilde{\nu}_{\text{inv}} \\ \tilde{\nu}_{\text{trans}} \end{cases}$$

(λ_{j+1})

Bull Math Biol (2014) 76:2249–2291
DOI 10.1007/s11538-014-9997-8

Society for Mathematical Biology

ORIGINAL ARTICLE

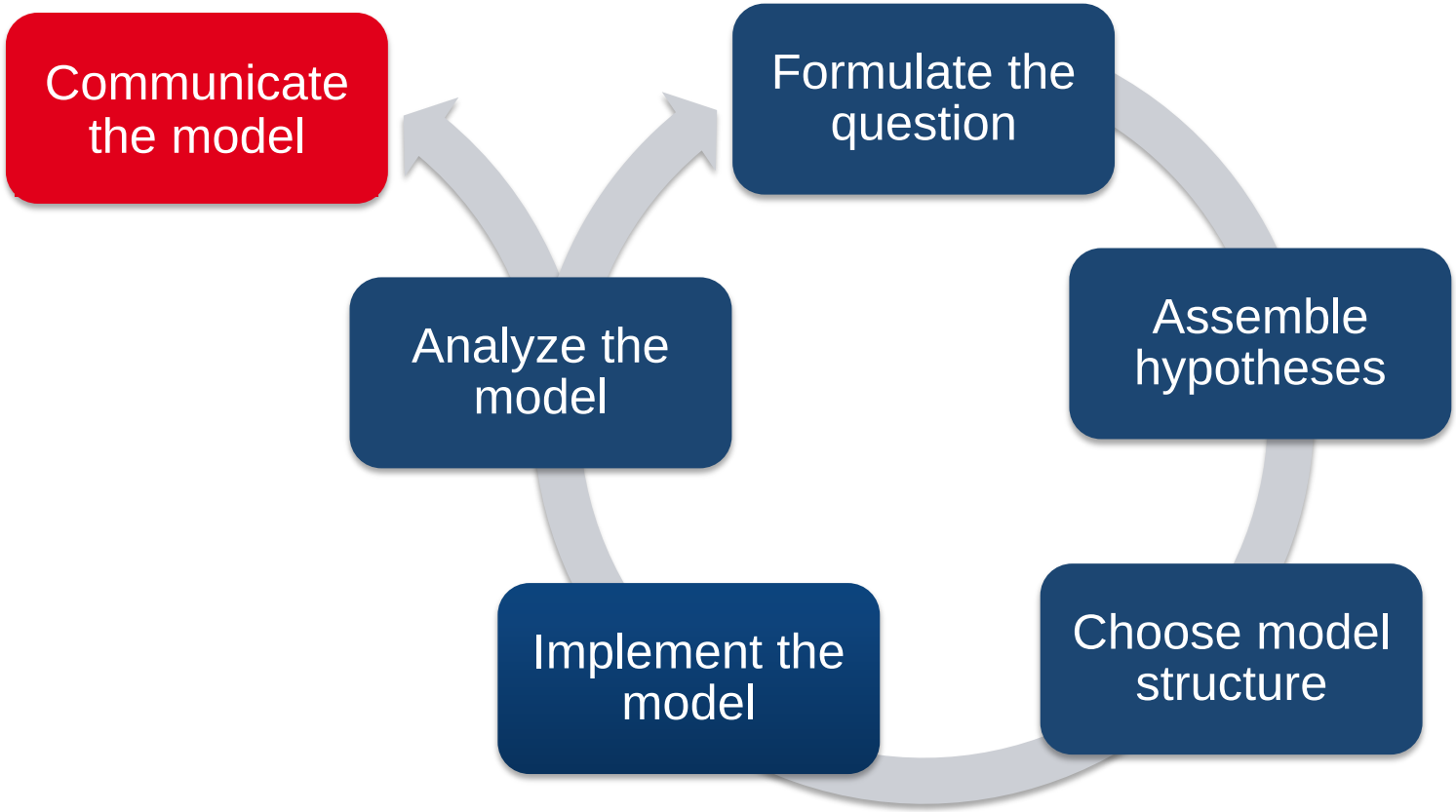
A Model for Genome Size Evolution

Stephan Fischer · Samuel Bernard ·
Guillaume Beslon · Carole Knibbe

Received: 15 November 2013 / Accepted: 15 July 2014 / Published online: 21 August 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract We present a model for genome size evolution that takes into account both local mutations such as small insertions and small deletions, and large chromosomal rearrangements such as duplications and large deletions. We introduce the possibility of undergoing several mutations within one generation. The model, albeit minimalist, reveals a non-trivial spontaneous dynamics of genome size: in the absence of selection, an arbitrary large part of genomes remains beneath a finite size, even for a duplication rate 2.6-fold higher than the rate of large deletions, and even if there is also a systematic bias toward small insertions compared to small deletions. Specifically, we show that

Le cycle de la modélisation individu-centrée



From [Railsback & Grimm, 2012]

Communiquer … mais quoi et à qui ?

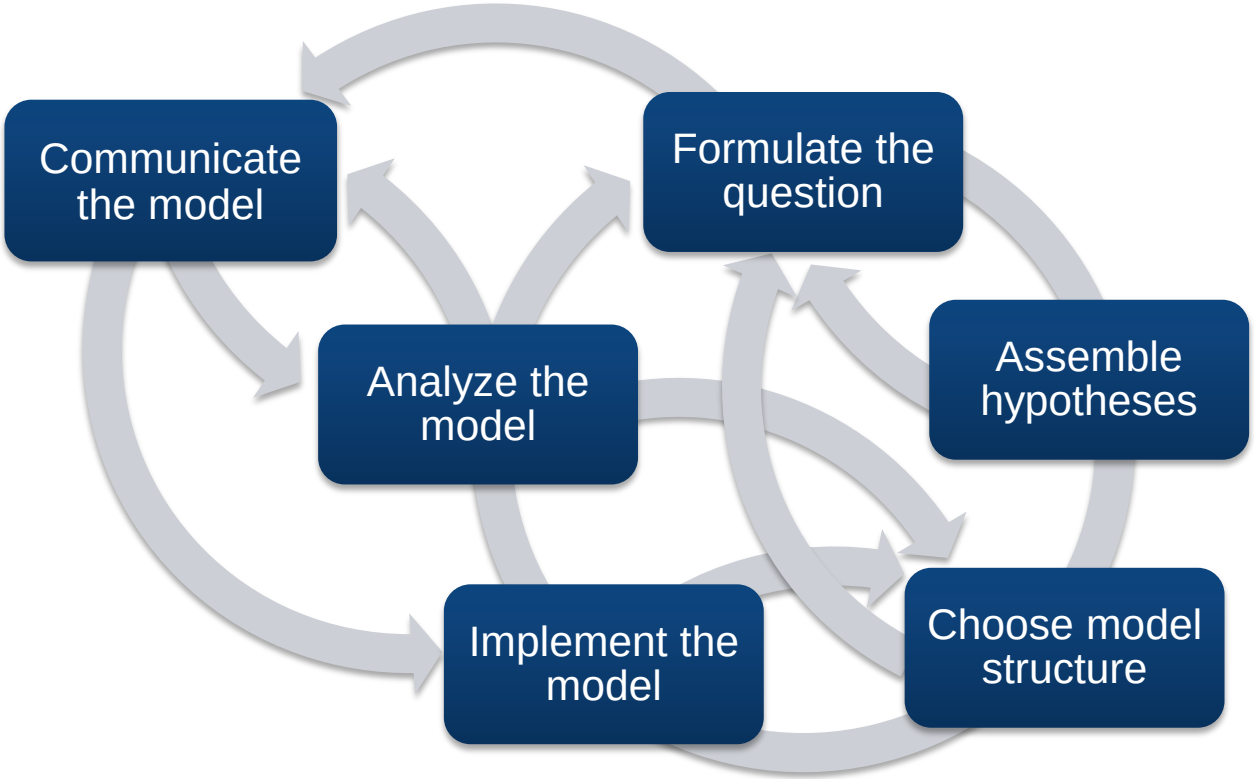
Message à l'attention des biologistes :

Si : (i) les génomes subissent des réarrangements de type délétion/duplication et (ii) le nombre et la taille moyenne de ces événements augmentent avec la taille du génome

- Alors : la variabilité mutationnelle d'une lignée dépend de la taille du chromosome, du nombre de gènes, de la taille des séquences intergéniques, de la taille des ARNs, …

- *Dans le modèle* la sélection indirecte pour un niveau de variabilité optimal ($F_{\wr}W$ ≈ 1) régule la taille du génome, les structures génomique (Knibbe *et al.*, Mol. Biol. Evol., 2007) et transcriptomiques (Parsons *et al.*, Alife XII, 2010, Beslon *et al.*, BioSystems, 2010). Le phénomène est robuste face à la sélection et aux biais mutationnels modérés (Fischer *et al.*, Bull. Math. Biol, 2014).

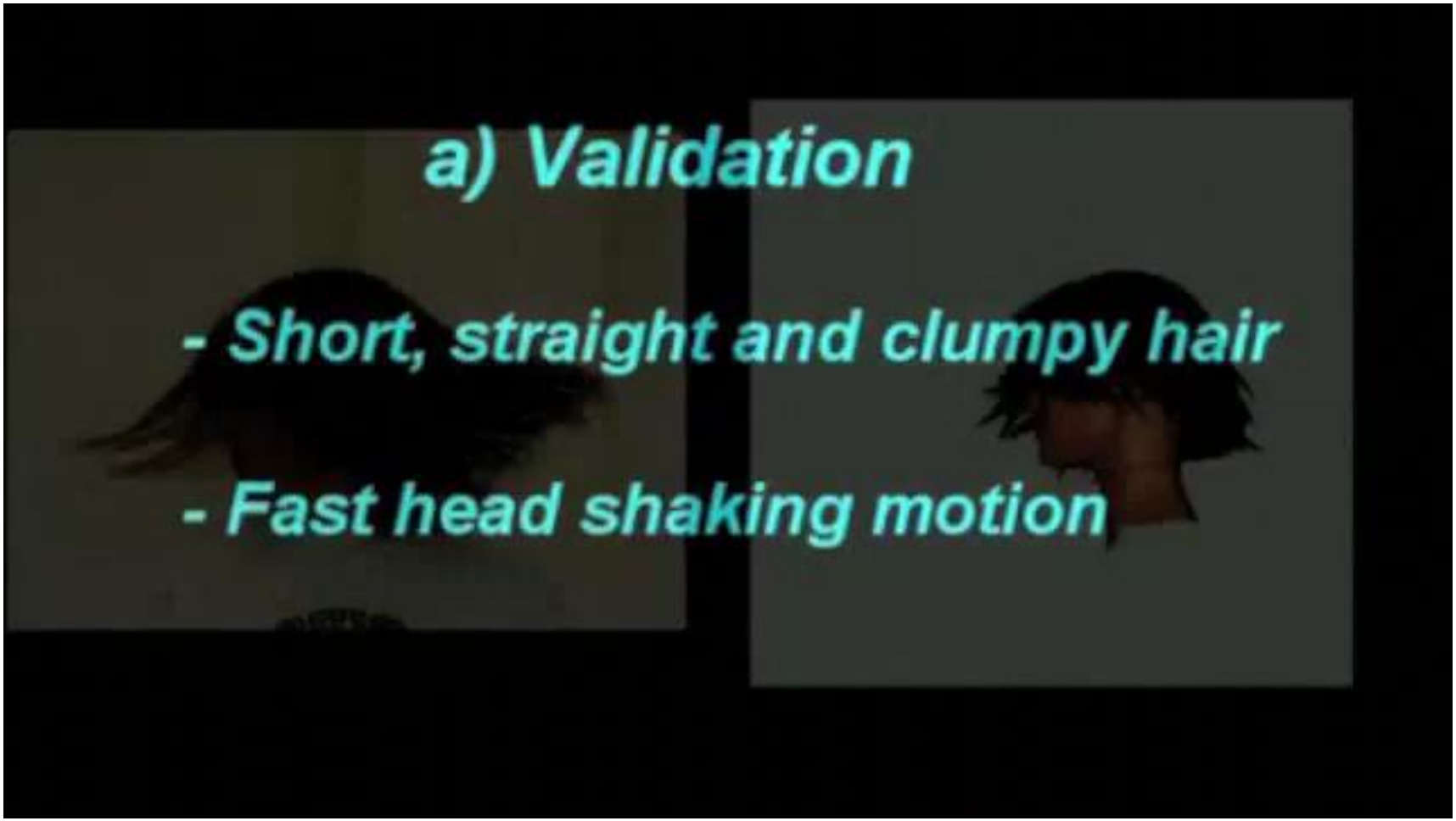
Conclusion (1) : ce que je ne vous ai pas dit …



From [Railsback & Grimm, 2012]

- Il est fondamental de distinguer deux phases dans la modélisation
  - La phase d'appropriation (co-construction mentale de la théorie et du modèle, voire de la question) → Le modèle est encore flou et mouvant
  - La phase d'exploitation → Le modèle est figé pour permettre une expérimentation rigoureuse

Conclusion (2) : la question à laquelle on ne coupe jamais …



[Bertails et al., 2006]

Conclusion (2) : la question à laquelle on ne coupe jamais …

- Peut-on valider et vérifier un modèle ?
  - Non et non mais pas pour les mêmes raisons
- Vérification du modèle
  - Le modèle est conforme à mon discours *ante* et me permet donc de construire un discours *post*
  - Généralement impossible pour un IBM mais on peut s'en approcher …
  - « Bonnes pratiques » de l'outil de modélisation
- Validation du modèle
  - Le modèle est conforme à l'objet (et à la question) ; il se comporte bien comme lui (dans le contexte de la question)
  - Tout dépend de la question : impossible dans un contexte de découverte, on peut s'en approcher dans un contexte de prédiction …
- Ne jamais confondre le modèle et son instanciation …

"Remember that all models are wrong; the practical question is how wrong do they have to be not to be useful" [G. Box]

Conclusion (3) : et donc, quel est le meilleur modèle ?

- Les modèles sont des instruments personnels
  - Le meilleur est donc le votre (pardon, le mien !)
  - Mais il ne pourra pas s'attaquer à n'importe quelle question !
  - En pratique il y a co-détermination des modèles et des questions …
- Quels est l'intérêt des modèles individu-centrés ?
  - [Grimm, 1999] : motivation pragmatique vs. motivation paradigmatique
    → Bonne gestion de l'hétérogénéïté des agents
  - Les motivations sont souvent plus frustres : les IBM …
    → Sont faciles à construire, à manipuler, à modifier … trop faciles ?
    → Peuvent modéliser des phénomènes inconnus (capacité à surprendre)
    → Utilisent l'ontologie du domaine et permettent donc une bonne communication interdisciplinaire (faciles à décrire/expliquer)
    → Sont plaisants ("look & feel") … trop plaisants ?
- En pratique il ne sert à rien d'opposer les modèles ; il vaut mieux rechercher des complémentarités …

Conclusion (4) : et donc, y a-t-il une méthodologie efficace ?

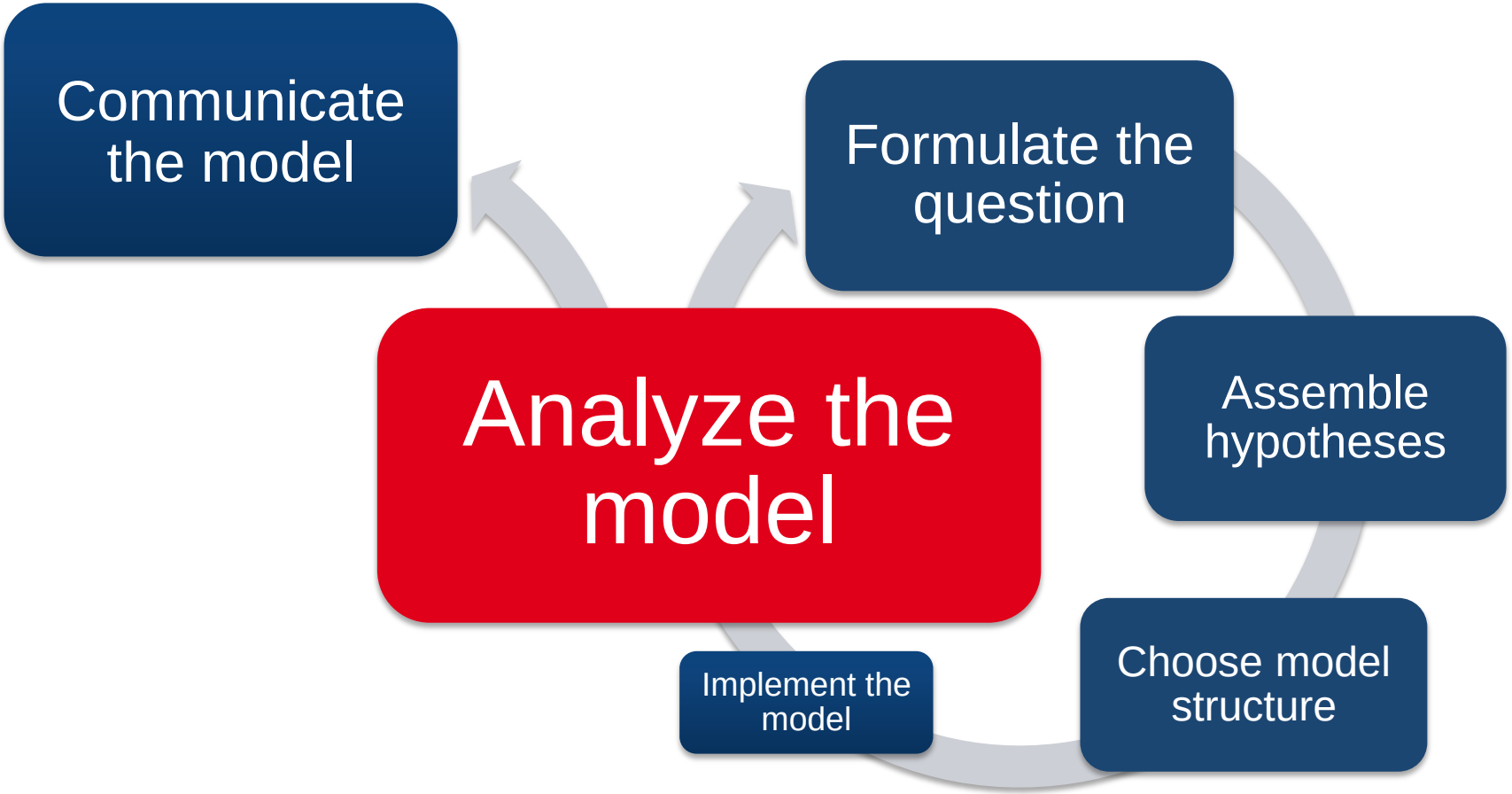
- Définitivement non ; chaque cas est particulier …
  - La modélisation est un art …
  - Un faussaire n'est pas un artiste, même si il est très bon …
    → L'apprentissage de la modélisation ne peut pas se faire en reprenant des modèles connus
- Mais il existe quelques points clés
  - Partir d'une bonne question (c.a.d. une question qui intéresse quelqu'un !)
  - Acquérir une très bonne maîtrise des outils de modélisation
  - Conduire les expériences rigoureusement

    "*Avoid the temptation to run tomorrow's computer simulations before yesterday's has been fully understood*" [miller, 1995]

  - Utiliser des modèles complémentaires, re-mathématiser les modèles
  - Confronter les résultats avec les spécialistes du domaine

    "The burden of proof is on us to explain our results to biologists in their own language and in their own journals" [miller, 1995]

Conclusion (4) : et donc, y a-t-il une méthodologie efficace ?



Bibliographie

Starfield, A. M., Smith, K. A., Bleloch, A. L. (1994) How to model it, problem solving for the computer age. Interaction book company, 208 p.

Grimm, V. (1999). Ten years of individual-based modelling in ecology : what have we learned and what could we learn in the future ? *Ecological Modelling*, 115:129–148.

Bonabeau, E. (2002). Agent-based modeling : Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 99(suppl. 3):7280–7287.

Bankes, S. (2002). Agent-based modeling : A revolution ? *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 99:7199–7200.

Grimm,V., and Railsback, S. F. (2005) Individual-based Modeling and Ecology, Princeton University Press, 428 p.

Raislback, S. F. and Grimm, V. (2012). Agent-Based and Individual-Based Modeling, a practical introduction. Princeton University Press, 329 p.

Hindré, T., Knibbe, C., Beslon, G., Schneider, D. (2012). New insights into bacterial adaptation through *in vivo* and *in silico* experimental evolution. *Nature Reviews Microbiology*, 10(5):352-365.

Varennes, F. and Silberstein, M. (Eds) (2013). Modéliser & Simuler. Epistémologie et pratique de la modélisation et de la simulation. Editions Matériologiques, 974 p.

La poésie ça sert surtout à compliquer les choses, non ?
Mais peut-être qu'en les compliquant on les comprend mieux.
Et en les comprenant on les simplifie,
au bout du compte.

Fred Vargas "Dans les bois éternels"
(ed. Viviane Hamy, 2006)