

Sébastien Clément

Biologiste

Équipe de génétique forestière (Jean Beaulieu)

Centre de foresterie des Laurentides

[Sebastien.Clement@RNCAN-NRCAN.gc.ca](mailto:Sebastien.Clement@RNCAN-NRCAN.gc.ca)

(418) 648-4886

Projet:

- Identification de gènes liés à la formation du bois chez les conifères (avec Arborea)

Tâches:

- Développement de bases de données:
  - Découverte de SNPs (TreeSNPs)
  - Données phénotypiques (PhenoTree)
  - Données de génotypage
- Développement de scripts (Unix) pour l'analyse et le formatage des données





## Références SQL intéressantes:

- Visual Quickstart Guide, SQL, 2<sup>nd</sup> edition, 461 p., Chris Fehily, 2005, Peachpit Press, Berkeley.

- [http://www.w3schools.com/sql/sql\\_quickref.asp](http://www.w3schools.com/sql/sql_quickref.asp)

- <http://www.1keydata.com/sql/sql.html>

- <http://www.digitalwidgets.com/Docs/SQLQuickReference.PDF>

## Cours suggéré:

UL:

IFT 2004 - Modèles et langages des bases de données

## Plan du cours

### 1. Introduction

- SGBD
- SQL

### 2. Pourquoi utiliser un SGBD ?

### 3. Comment créer une base de données

- besoins
- modélisation
- SQL

### 4. Interroger une base de données

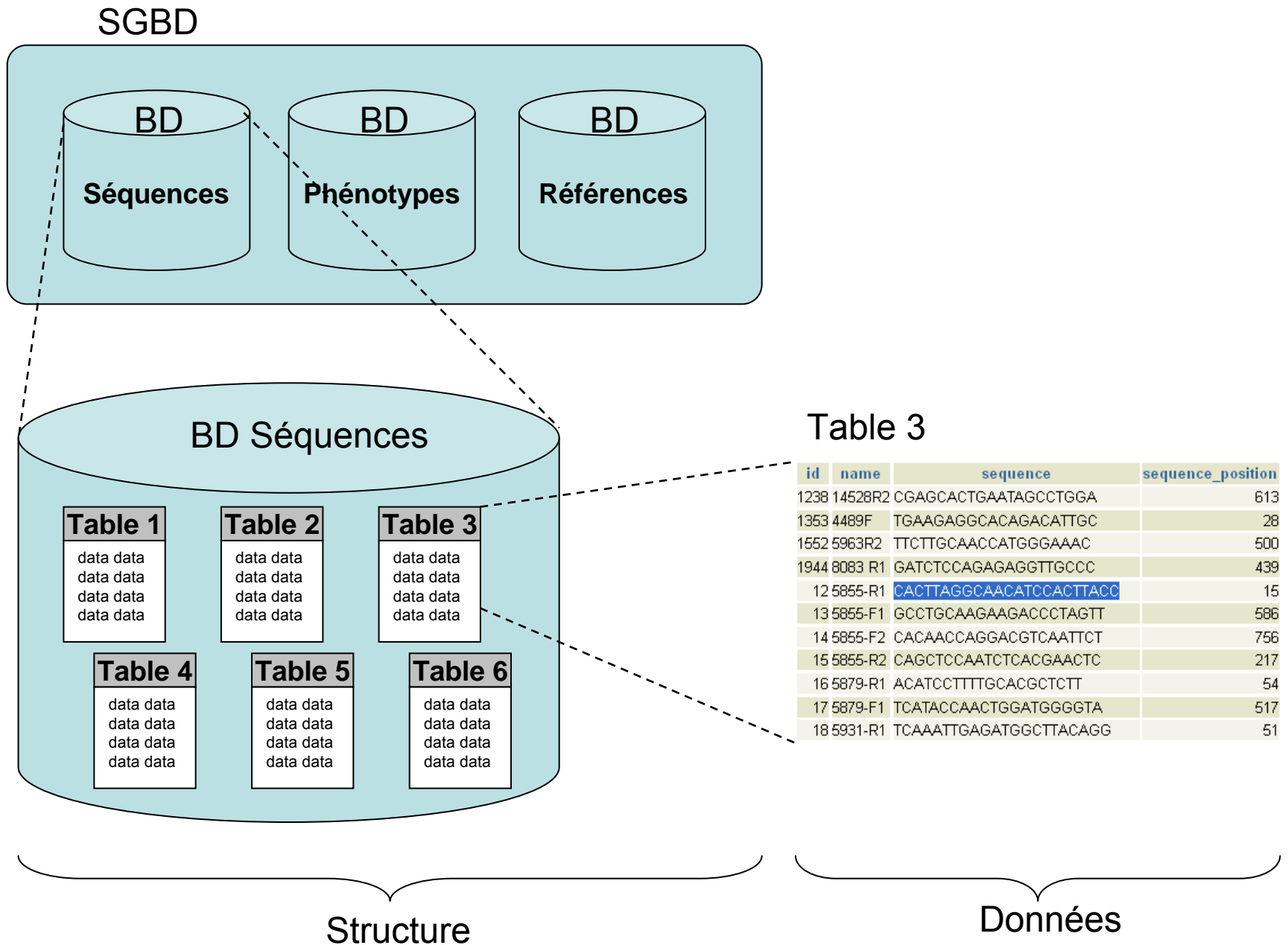


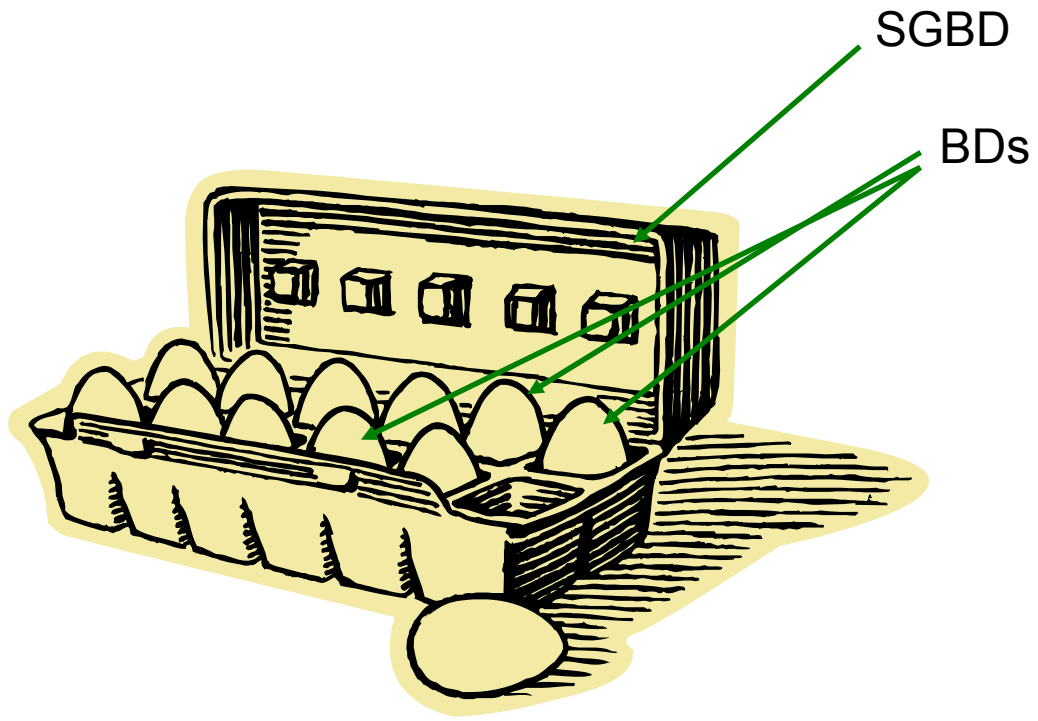
Théorie

Pratique

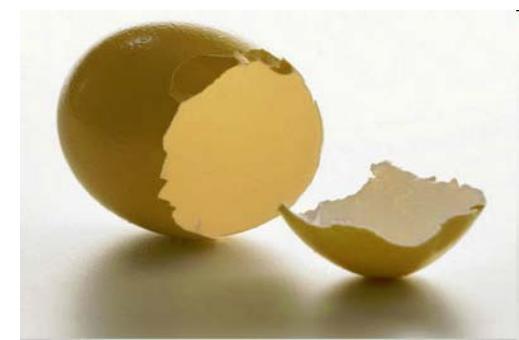
## SGBD

- Système de gestion de bases de données (anglais: *DBMS* ou *RDBMS*)
- Exemples:
  - Access
  - MS SQL Server
  - Oracle
  - MySQL
  - PostgreSQL
- Caractéristiques:
  - Langage commun: SQL
  - Système sécurisé
  - Opérations concurrentes
  - Intégrité
  - Catalogue
  - Sauvegarde et récupération
  - Vues

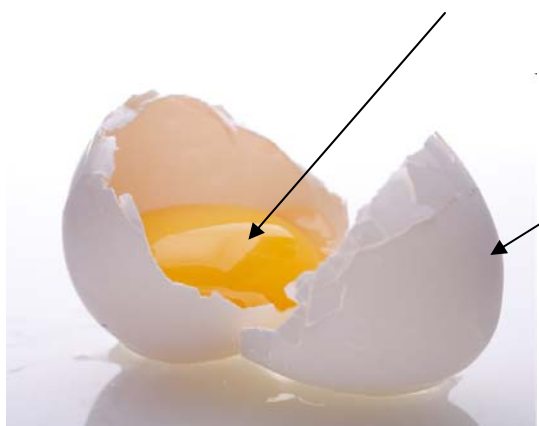




BD  
sans  
données

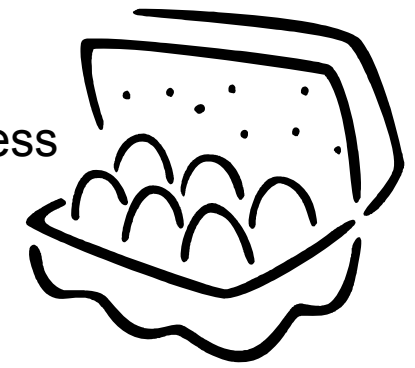


Données



Structure

MS Access



## BD - exemples

- Banques de données biologiques publiques
  - Genbank
  - TAIR
  - Flybase
  - etc.
- De façon générale - tout ce qui gère des données:
  - Banques – ex: transactions aux guichets
  - Sites d'achat en ligne (Ebay, Amazon, Expedia, Hotwire, etc.)
  - GIS
  - etc.



## SQL

- *Structured Query Language*
- Standard ISO
- Commun aux SGBD, mais certains ont des spécificités
  
- Plus que de la recherche de données:
  - DDL – Data Definition Language
  - DML – Data Manipulation Language
  
- Avantages:
  - Simple
  - Rapide
  - Non-procédural
  
- Exemple:

```
SELECT champ1, champ5 FROM ma_table;
```

## Pourquoi utiliser un SGBD ?

	Fichiers à plat, Excel	SGBD
+	<ul style="list-style-type: none"><li>•Rapide à mettre sur pied</li><li>•Facile d'utilisation (Excel)</li></ul>	<ul style="list-style-type: none"><li>•Données regroupées</li><li>•Accès concurrent</li><li>•Accès sécurisé</li><li>•Sauvegarde, restauration</li><li>•Contrôle d'intégrité (validation)</li><li>•Mise-à-jour des données</li><li>•Recherche facile (SQL) et rapide (index)</li><li>•Vues</li></ul>
-	<ul style="list-style-type: none"><li>•Capacité limitée (Excel)</li><li>•Aucun accès concurrent</li><li>•Aucun contrôle d'intégrité</li><li>•Accès non-sécurisé</li><li>•Redondance et multiples-versions des données</li><li>•Jointures ???</li></ul>	<ul style="list-style-type: none"><li>•Plus long à monter</li><li>•Doit être bien planifié</li><li>•Performance de calcul</li><li>•Taille</li></ul>

## Comment choisir un SGBD ?

	<b>Propriétaires</b>	<b>Libres</b>
<b>+</b>	<ul style="list-style-type: none"><li>•Meilleur support technique*</li><li>•Documentation</li></ul>	<ul style="list-style-type: none"><li>•Gratuit</li><li>•Souvent aussi performant</li></ul>
<b>-</b>	<ul style="list-style-type: none"><li>•Prix**</li><li>•Lourdeur</li></ul>	<ul style="list-style-type: none"><li>•Support technique souvent déficient (forums...)</li><li>•Complexité ?</li></ul>
	<ul style="list-style-type: none"><li>•MS Access</li><li>•MS SQL Server</li><li>•Oracle</li><li>•Sybase</li></ul>	<ul style="list-style-type: none"><li>•MySQL</li><li>•PostgreSQL</li><li>•SQLite</li></ul>

\*Souvent coûts en sus du coût d'achat

\*\*SQL Server: 9 500 à 36 700 \$ CAN (selon version), license de 2 ans

## Interfaces-utilisateur: ligne de commande (SQL) ou interface graphique ?

## SQL+ (Oracle)

```

C:\Temp>
C:\Temp>
C:\Temp>
C:\Temp>
C:\Temp>type login.sql
set echo off
set heading off
set feedback off
prompt
prompt "Welcome to SQL*Plus ..."
select 'Database = '||instance_name from v$instance;
select 'UserName = '||username from user_users;
prompt
set heading on
C:\Temp>
C:\Temp>
C:\Temp>sqlplus bert/bert

SQL*Plus: Release 10.2.0.3.0 - Production on Sun Jan 27 12:18:26 2008
Copyright (c) 1982, 2006, Oracle. All Rights Reserved.

Connected to:
Oracle Database 10g Enterprise Edition R
With the Partitioning, OLAP and Data Min

"Welcome to SQL*Plus ..."

Database = orxp10
UserName = BERT
SQL>
SQL>
SQL>
SQL>
SQL>

```

psql (PostgreSQL)

```

c:\ Invite de commandes - psql -U cleseb -d phenotree_development -h "132.156.208.45"

```

Schéma	Nom	Type	Propriétaire
public	alternative_germplasm_names	table	cleseb
public	bad_rings	table	cleseb
public	branch_measures	table	cleseb
public	bud_development_codes	table	cleseb
public	bud_growth	table	cleseb
public	canadian_forest_ecozones	table	cleseb
public	canadian_forest_sections	table	cleseb
public	chemical_measures	table	cleseb
public	coded_observations_on_germplasms	table	cleseb
public	comments	table	cleseb
public	comments_on_germplasms	table	cleseb
public	containers	table	cleseb
public	cross_objectives	table	cleseb
public	data_source_references	table	cleseb
public	data_sources	table	cleseb
public	employees	table	cleseb
public	employees_in_teams	table	cleseb
public	families	table	cleseb
public	family_provenances	table	cleseb
public	field_descriptions	table	cleseb
public	general_growth	table	cleseb

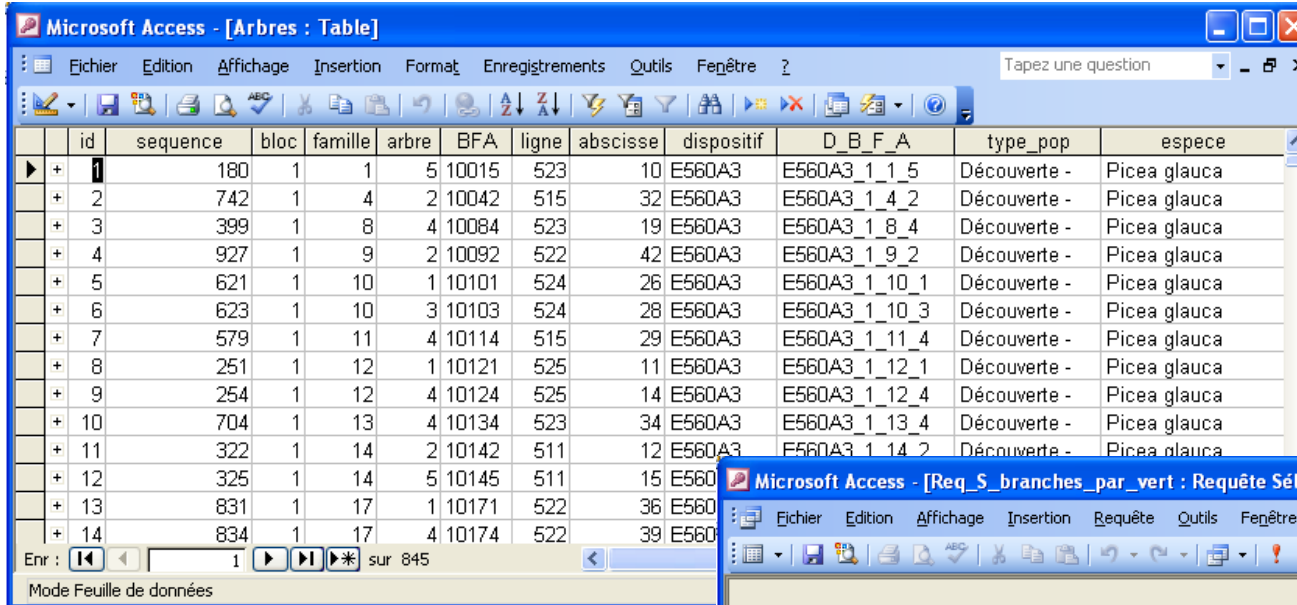
```

phenotree_development=> SELECT * FROM employees;

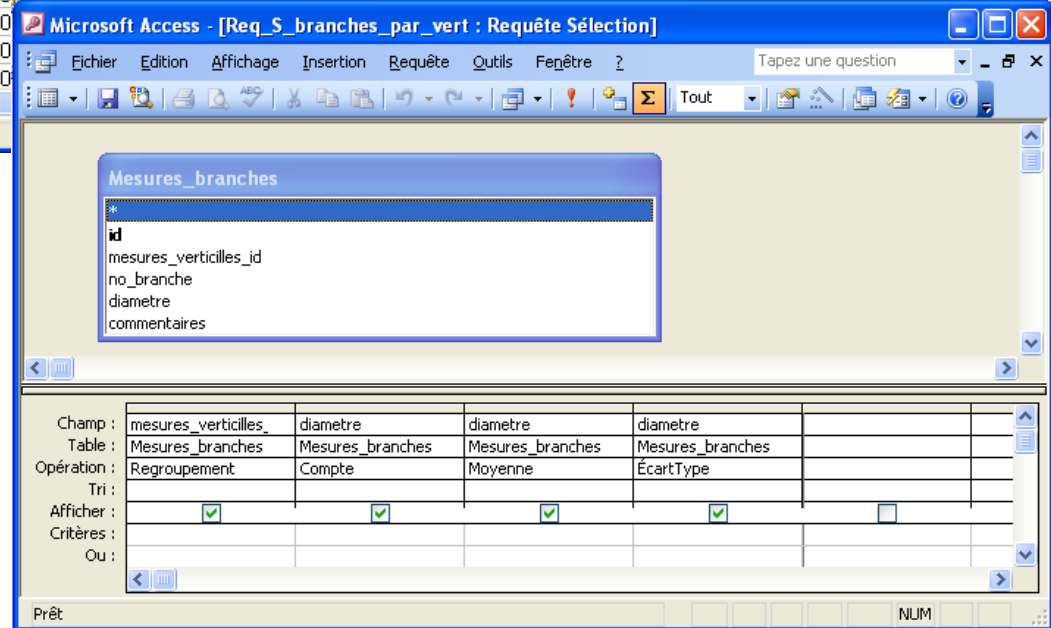
```

## Interfaces-utilisateur: ligne de commande (SQL) ou interface graphique ?

## Access – formulaire d'édition des enregistrements



	id	sequence	bloc	famille	arbre	BFA	ligne	abscisse	dispositif	D_B_F_A	type_pop	espece
+	1	180	1	1	5	10015	523	10	E560A3	E560A3_1_1_5	Découverte -	Picea glauca
+	2	742	1	4	2	10042	515	32	E560A3	E560A3_1_4_2	Découverte -	Picea glauca
+	3	399	1	8	4	10064	523	19	E560A3	E560A3_1_8_4	Découverte -	Picea glauca
+	4	927	1	9	2	10092	522	42	E560A3	E560A3_1_9_2	Découverte -	Picea glauca
+	5	621	1	10	1	10101	524	26	E560A3	E560A3_1_10_1	Découverte -	Picea glauca
+	6	623	1	10	3	10103	524	28	E560A3	E560A3_1_10_3	Découverte -	Picea glauca
+	7	579	1	11	4	10114	515	29	E560A3	E560A3_1_11_4	Découverte -	Picea glauca
+	8	251	1	12	1	10121	525	11	E560A3	E560A3_1_12_1	Découverte -	Picea glauca
+	9	254	1	12	4	10124	525	14	E560A3	E560A3_1_12_4	Découverte -	Picea glauca
+	10	704	1	13	4	10134	523	34	E560A3	E560A3_1_13_4	Découverte -	Picea glauca
+	11	322	1	14	2	10142	511	12	E560A3	E560A3_1_14_2	Découverte -	Picea glauca
+	12	325	1	14	5	10145	511	15	E560			
+	13	831	1	17	1	10171	522	36	E560			
+	14	834	1	17	4	10174	522	39	E560			



Microsoft Access - [Req\_S\_branches\_par\_vert : Requête Sélection]

Mesures\_branches

```

*
id
mesures_verticilles_id
no_branche
diametre
commentaires

```

Champ :	mesures_verticilles_	diametre	diametre	diametre
Table :	Mesures_branches	Mesures_branches	Mesures_branches	Mesures_branches
Opération :	Regroupement	Compte	Moyenne	ÉcartType
Tri :				
Afficher :	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Critères :				
Ou :				

Prêt

## Access – constructeur de requêtes

## Créer une base de données: besoins, modélisation et création SQL

### Besoins de l'équipe

- Quel est le but ?
- Quelles données ?
- Utilisation concurrente ?
- Court-terme vs long-terme ?
- Rôles des utilisateurs ?
  - administrateur de la bd (DBA)
  - administrateur des données (DA)
  - utilisateur en lecture seule



Créer une base de données: besoins, modélisation et création SQL

## ***La modélisation des données***

Modèle de données:

**« Description des données (1), des liens entre elles (2) et des contraintes sur les données (3), dans une organisation. »**

## 1. Description des données

*Mon équipe de recherche a:*

- *ADN (cDNA, gDNA)*
- *amorces*
- *résultats PCR*

ADNs

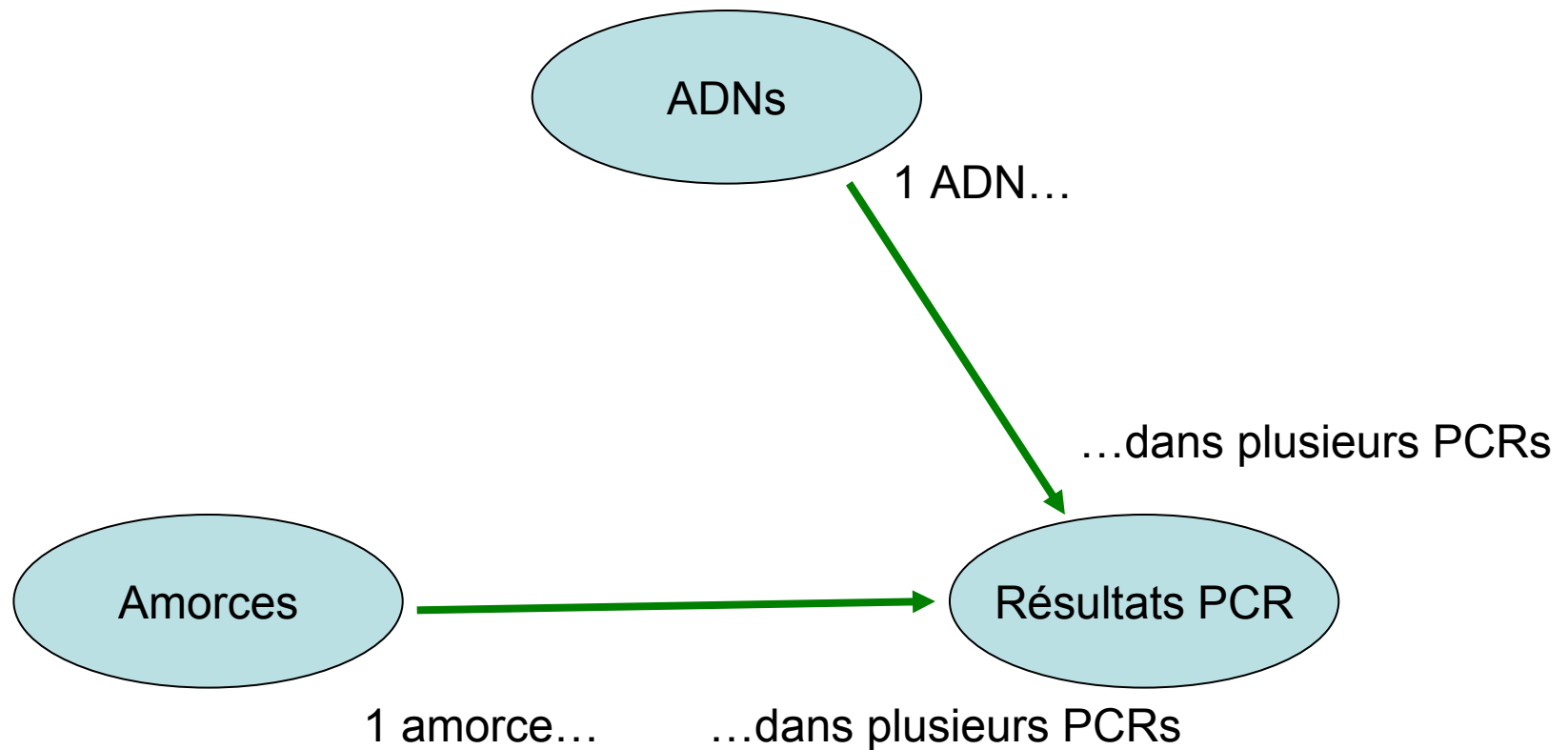
Amorces

Résultats PCR

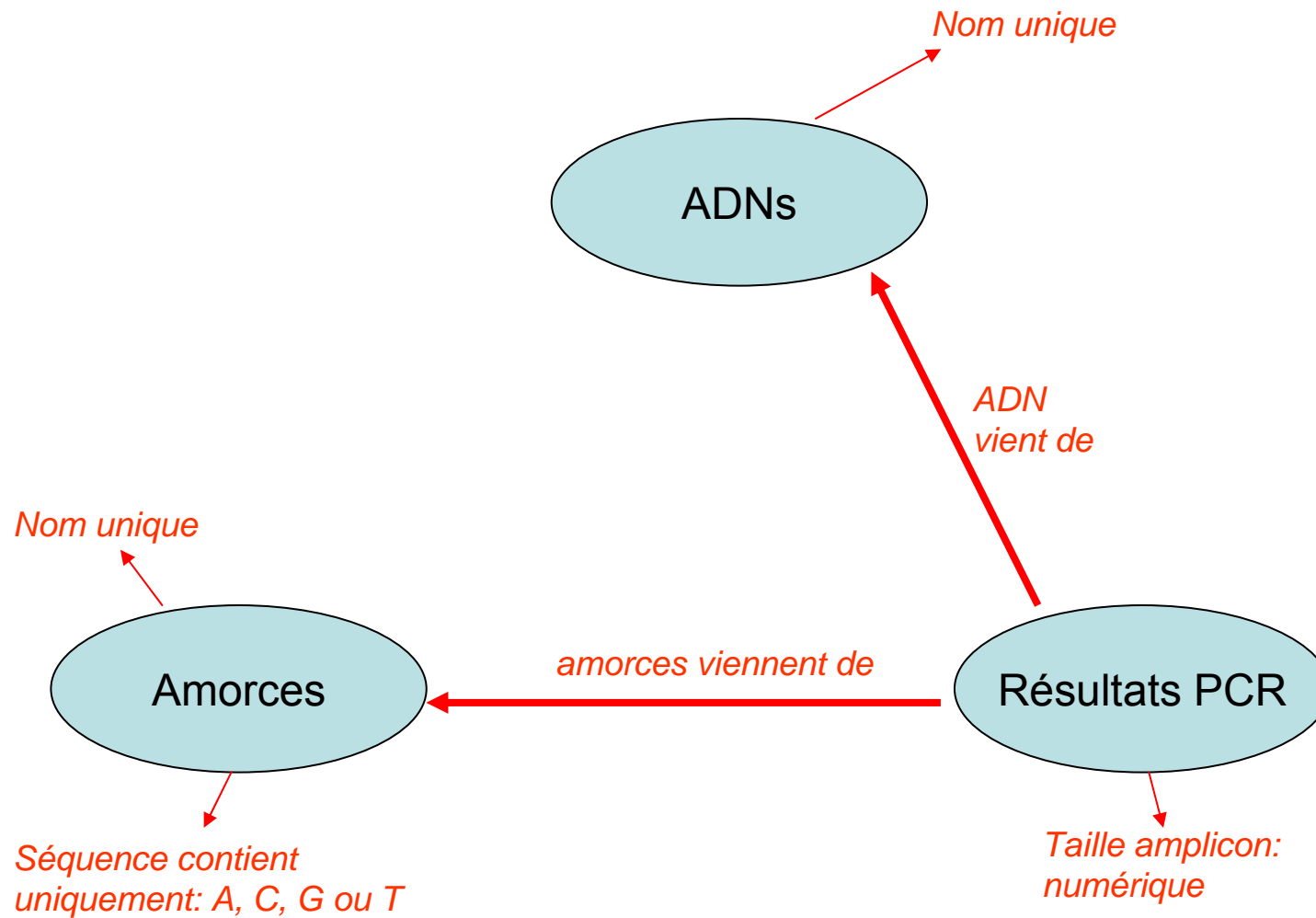


## 2. Liens entre les données

• **Les amorces et les ADN sont mis en commun pour faire un PCR**



### 3. Contraintes sur les données



## 1. Description des données

*Mon équipe de recherche a:*

- **ADN**
- **amorces**
- **résultats PCR**

### ADNs

nom	espèce	type
at_312	A. thaliana	génomique
pt_221s	P. taeda	cDNA
pg_S352	P. glauca	cDNA
ADN_K33	P. aeruginosa	plasmidique
ce_ttY14	C. elegans	cDNA

### Amorces

nom	séquence	sens
at39234F	ACACACAACAACACAATTCG	F
at39234R	CATACATATCATACTATCAACTA	R
pt00324F	GCCTGCAAGAAGACCCTAGTT	F
pt00324R	CACTTAGGCAACATCCACTTACC	R

### PCRs

adn_nom	amorce_f	amorce_r	programme	melange	taille_amplicon
at_312	at39234F	at39234R	Prog0077	Mix321	
at_312	at39234F	at39234R	Prog0002	Mix321	1400
ADN_K33	K33_001F	K33_001R	Prog0012	Mix217	
ADN_K33	K33_001F	K33_001R	Prog0012	Mix218	450

## 2. Liens entre les données

**Les amorces et les ADN  
sont mis en commun  
pour faire un PCR**

ADNs

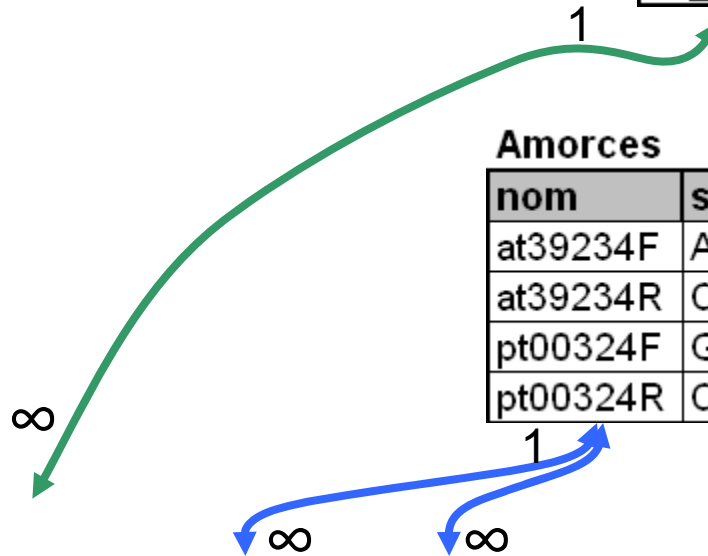
nom	espèce	type
at_312	A. thaliana	génomique
pt_221s	P. taeda	cDNA
pg_S352	P. glauca	cDNA
ADN_K33	P. aeruginosa	plasmidique
ce_ttY14	C. elegans	cDNA

Amorces

nom	séquence	sens
at39234F	ACACACAACAACACAATTTCG	F
at39234R	CATACATATCATACTATCAACTA	R
pt00324F	GCCTGCAAGAAGACCCTAGTT	F
pt00324R	CACTTAGGCAACATCCACTTACC	R

PCRs

adn_nom	amorce_f	amorce_r	programme	melange	taille_amplicon
at_312	at39234F	at39234R	Prog0077	Mix321	
at_312	at39234F	at39234R	Prog0002	Mix321	1400
ADN_K33	K33_001F	K33_001R	Prog0012	Mix217	
ADN_K33	K33_001F	K33_001R	Prog0012	Mix218	450



## 3. Contraintes sur les données

ADNs

nom	espèce	type
at_312	A. thaliana	génomique
pt_221s	P. taeda	cDNA
pg_S352	P. glauca	cDNA
ADN_K33	P. aeruginosa	plasmidique
ce_ttY14	C. elegans	cDNA

*Unique*

Amorces

nom	séquence	sens
at39234F	ACACACAACAACACAATTCG	F
at39234R	CATACATATCATACTATCAACTA	R
pt00324F	GCCTGCAAGAAGACCCTAGTT	F
pt00324R	CACTTAGGCAACATCCACTTACC	R

*Doit provenir de**A, C, G ou T*

PCRs

adn_nom	amorce_f	amorce_r	programme	melange	taille_amplicon
at_312	at39234F	at39234R	Prog0077	Mix321	
at_312	at39234F	at39234R	Prog0002	Mix321	1400
ADN_K33	K33_001F	K33_001R	Prog0012	Mix217	
ADN_K33	K33_001F	K33_001R	Prog0012	Mix218	450

*Numérique*

### 3. Contraintes sur les données



#### Primaire (PK):

identifie une ligne de façon unique



#### Étrangère (FK):

réfère à la clé primaire d'une autre table

#### ADNs

nom	espèce	type
at_312	A. thaliana	génomique
pt_221s	P. taeda	cDNA
pg_S352	P. glauca	cDNA
ADN_K33	P. aeruginosa	plasmidique
ce_ttY14	C. elegans	cDNA

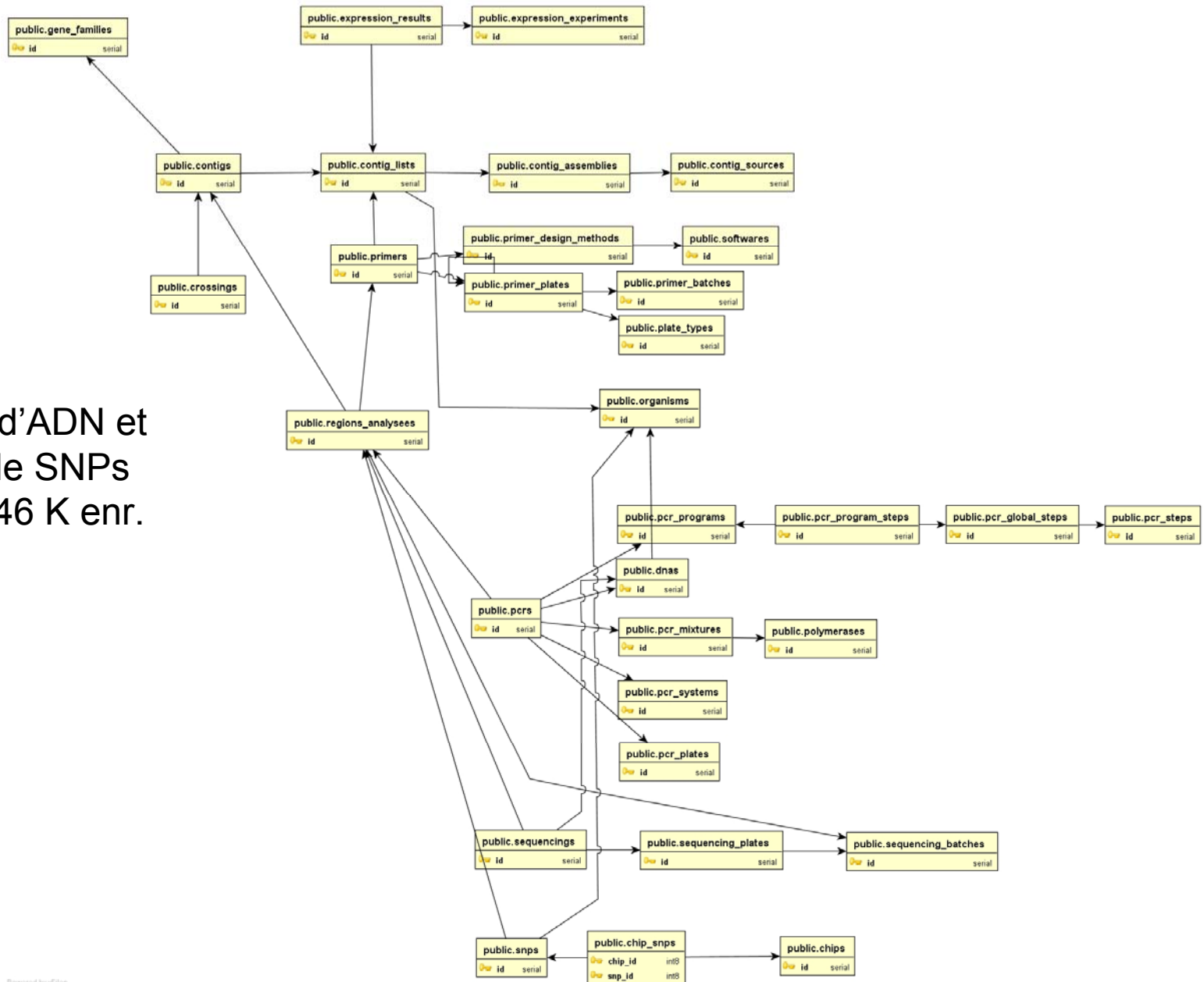
#### Amorces

nom	séquence	sens
at39234F	ACACACAACAACACAATTTCG	F
at39234R	CATACATATCATACTATCAACTA	R
pt00324F	GCCTGCAAGAAGACCCTAGTT	F
pt00324R	CACTTAGGCAACATCCACTTACC	R

#### PCRs

id	adn_nom	amorce_f	amorce_r	programme	melange	taille_amplicon
1	at_312	at39234F	at39234R	Prog0077	Mix321	
2	at_312	at39234F	at39234R	Prog0002	Mix321	1400
3	ADN_K33	K33_001F	K33_001R	Prog0012	Mix217	
4	ADN_K33	K33_001F	K33_001R	Prog0012	Mix218	450

?

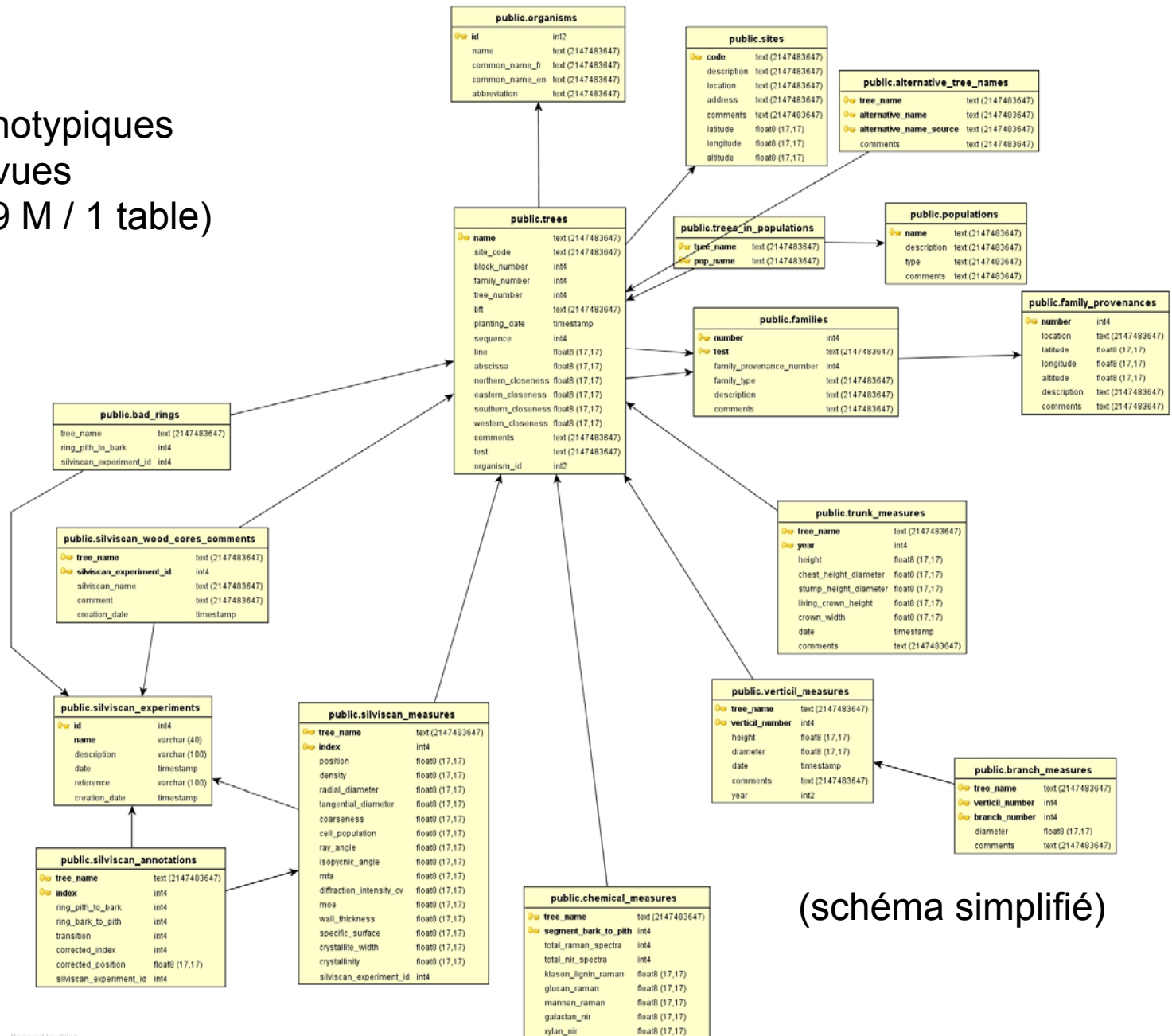


## TreeSNPs

- Séquences d'ADN et découverte de SNPs
- 38 tables, 446 K enr.

## PhenoTree

- Données phénotypiques
- 61 tables, 26 vues
- 7.4 M enr. (5.9 M / 1 table)



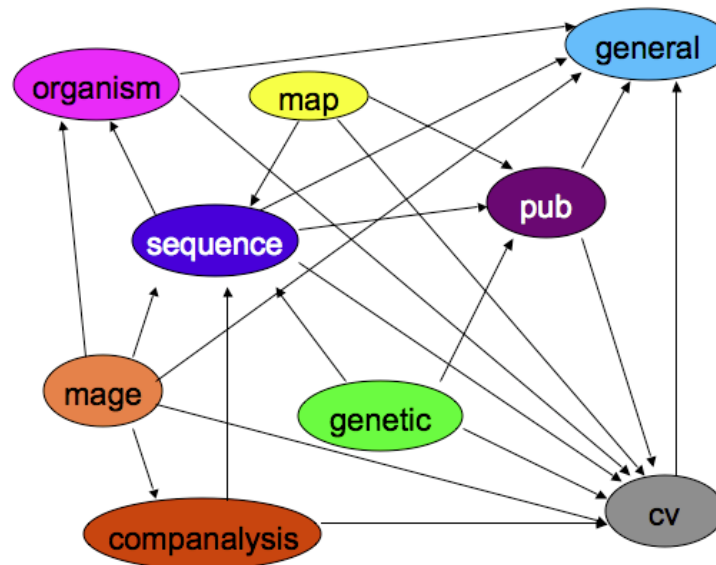
(schéma simplifié)



Schémas déjà existants sur le Web?

Génomique

Chado - <http://gmod.org/wiki/Chado>



+	<ul style="list-style-type: none"><li>•Générique</li><li>•Plusieurs sont gratuits</li><li>•Déjà utilisé et amélioré par d'autres</li><li>•Interfaces graphiques</li></ul>
-	<ul style="list-style-type: none"><li>•Générique !</li><li>•Maîtrise du schéma</li><li>•« Overkill » ?</li></ul>

BioSQL - [http://www.biosql.org/wiki/Main\\_Page](http://www.biosql.org/wiki/Main_Page)

La création de la base de données en SQL

Serveur de bases de données  
SGBD: **PostgreSQL**

Site: <http://dbs.arborea.ulaval.ca/phpPgAdmin/>  
Interface Web: PhpPgAdmin

Base de données: **ibis\_db**

The image shows a screenshot of the phpPgAdmin 4.2.3 web interface. On the left, a sidebar menu under 'Servers' has 'PostgreSQL' highlighted with a red box. A green arrow points from this box to the main content area. The main area displays 'phpPgAdmin 4.2.3 (PHP 5.2.13)' and a 'Login to PostgreSQL' form. The form includes fields for 'Username' (containing 'ibis01') and 'Password' (masked with dots), and a 'Login' button. The interface also shows a language dropdown set to 'English' and a list of links: 'phpPgAdmin Homepage', 'PostgreSQL Homepage', 'Report a Bug', and 'View online FAQ'.

## Partie pratique 1

### Créer une base de données

Voir le document  
`commandes_sql.pdf`

## Partie pratique 2

### Interroger une base de données

## Interroger une base de données – requête SQL (DML)

À quoi servent les requêtes SQL ?

- A) Aller chercher de l'info de façon très précise (critères)
- B) Joindre de l'info répartie dans plusieurs tables
- C) Effectuer des calculs par ligne, ou sur plusieurs lignes

Avantages:

- Langage simple
- Dynamique (mise-à-jour)
- Soumis aux privilèges d'accès utilisateurs

## Interroger une base de données – requête SQL (DML)

Exemple A: Aller chercher de l'info de façon très précise (critères)

## Table amorces

nom	sequence	sens	adn_gabarit	pos_vs_gabarit	date_creation
at39234F	ACACACAACAACACAATTCG	F	at_312	43	2011-02-12
at39234R	CATACATATCATACTATCAACTA	R	at_312	512	2011-02-12
pt00324F	GCCTGCAAGAAGACCCTAGTT	F	pt_221s	52	2011-03-05
pt00324R	CACCTTAGGCAACATCCACTTACC	R	pt_221s	744	2011-03-05
K33_001F	CCAGATCTACTACATACTTACTAT	F	ADN_K33	23	2010-09-07
K33_001R	CAACATCTATCTACTATCTAT	R	ADN_K33	455	2010-09-07
K33_002F	CACACACACAATACGACACAT	F	ADN_K33	433	2010-09-07
K33_002R	CGGCGATTATTCGATTACGATCT	R	ADN_K33	971	2010-09-07

Sélection des colonnes

nom	date_creation
at39234F	2011-02-12
at39234R	2011-02-12
pt00324F	2011-03-05
pt00324R	2011-03-05
K33_001F	2010-09-07
K33_001R	2010-09-07
K33_002F	2010-09-07
K33_002R	2010-09-07

Sélection des lignes

nom	sequence	sens	adn_gabarit	pos_vs_gabarit	date_creation
at39234F	ACACACAACAACACAATTCG	F	at_312	43	2011-02-12
pt00324F	GCCTGCAAGAAGACCCTAGTT	F	pt_221s	52	2011-03-05
K33_001F	CCAGATCTACTACATACTTACTAT	F	ADN_K33	23	2010-09-07
K33_002F	CACACACACAATACGACACAT	F	ADN_K33	433	2010-09-07

## Interroger une base de données – requête SQL (DML)

## Exemple B: Joindre de l'info répartie dans plusieurs tables

Table ADNs

nom	espece	type
at_312	A. thaliana	Génomique
pt_221s	P. taeda	cDNA
pg_S352	P. glauca	cDNA
ADN_K33	P. aeruginosa	Plasmidique
ce ttY14	C. elegans	cDNA

Table amorces

nom	sequence	sens	adn_gabarit	pos_vs_gabarit	date_creation
at39234F	ACACACAACAACACAATTCG	F	at_312	43	2011-02-12
at39234R	CATACATATCATACTATCAACTA	R	at_312	512	2011-02-12
pt00324F	GCCTGCAAGAAGACCCTAGTT	F	pt_221s	52	2011-03-05
pt00324R	CACTTAGGCAACATCCACTTACC	R	pt_221s	744	2011-03-05
<K33_001F	CCAGATCTACTACATACTTACTAT	F	ADN_K33	23	2010-09-07
<K33_001R	CAACATCTATCTACTATCTAT	R	ADN_K33	455	2010-09-07
<K33_002F	CACACACACAATACGACACAT	F	ADN_K33	433	2010-09-07
<K33_002R	CGGCGATTATTCGATTACGATCT	R	ADN_K33	971	2010-09-07

ADNs et amorces correspondantes

nom	espece	type	nom	sequence
at_312	A. thaliana	Génomique	at39234R	CATACATATCATACTATCAACTA
at_312	A. thaliana	Génomique	at39234F	ACACACAACAACACAATTCG
pt_221s	P. taeda	cDNA	pt00324R	CACTTAGGCAACATCCACTTACC
pt_221s	P. taeda	cDNA	pt00324F	GCCTGCAAGAAGACCCTAGTT
pg_S352	P. glauca	cDNA	NULL	NULL
ADN_K33	P. aeruginosa	Plasmidique	K33_002R	CGGCGATTATTCGATTACGATCT
ADN_K33	P. aeruginosa	Plasmidique	K33_002F	CACACACACAATACGACACAT
ADN_K33	P. aeruginosa	Plasmidique	K33_001R	CAACATCTATCTACTATCTAT
ADN_K33	P. aeruginosa	Plasmidique	K33_001F	CCAGATCTACTACATACTTACTAT
ce ttY14	C. elegans	cDNA	NULL	NULL



## Interroger une base de données – requête SQL (DML)

## Exemple C: Effectuer des calculs par ligne, ou sur plusieurs lignes

## Table amorces

nom	sequence	sens	adn_gabarit	pos_vs_gabarit	date_creation
at39234F	ACACACAACAACACAATTTCG	F	at_312	43	2011-02-12
at39234R	CATACATATCATACTATCAACTA	R	at_312	512	2011-02-12
pt00324F	GCCTGCAAGAAGACCCTAGTT	F	pt_221s	52	2011-03-05
pt00324R	CACTTAGGCAACATCCACTTACC	R	pt_221s	744	2011-03-05
K33_001F	CCAGATCTACTACATACTTACTAT	F	ADN_K33	23	2010-09-07
K33_001R	CAACATCTATCTACTATCTAT	R	ADN_K33	455	2010-09-07
K33_002F	CACACACACAATACGACACAT	F	ADN_K33	433	2010-09-07
K33_002R	CGGCGATTATTCGATTACGATCT	R	ADN_K33	971	2010-09-07

Calcul par ligne

Calcul sur plusieurs lignes

nom	sequence	pct_gc	length
at39234F	ACACACAACAACACAATTTCG	40.0	20
at39234R	CATACATATCATACTATCAACTA	26.1	23
pt00324F	GCCTGCAAGAAGACCCTAGTT	52.4	21
pt00324R	CACTTAGGCAACATCCACTTACC	47.8	23
K33_001F	CCAGATCTACTACATACTTACTAT	33.3	24
K33_001R	CAACATCTATCTACTATCTAT	28.6	21
K33_002F	CACACACACAATACGACACAT	42.9	21
K33_002R	CGGCGATTATTCGATTACGATCT	43.5	23

sens	nb_amorces	avg_pos	min_pos	max_pos
F	4	138	23	433
R	4	671	455	971

Champs calculés

Voir le document  
`commandes_sql.pdf`

Discussions avancées:  
Transactions  
Normalisation

## Gestion des transactions: opérations concurrentes

Exemple sans gestion des transactions:

Temps	T <sub>1</sub>	T <sub>2</sub>	solde <sub>x</sub>
t <sub>1</sub>		début_transaction	100
t <sub>2</sub>	début_transaction	lire(solde <sub>x</sub> ) → 100	100
t <sub>3</sub>	lire(solde <sub>x</sub> ) → 100	solde <sub>x</sub> = solde <sub>x</sub> + 100 → 200	100
t <sub>4</sub>	solde <sub>x</sub> = solde <sub>x</sub> - 10 → 90	écrire(solde <sub>x</sub> ) → 200	200
t <sub>5</sub>	écrire(solde <sub>x</sub> ) → 90	validation	90
t <sub>6</sub>	validation		90

Normalisation: 1. **non-normalisée***Autant de colonnes que d'appartenance à des projets!***Employés**

#	Prénom	Nom de famille	Local	Projet1	Projet2	Projet3	Projet4	Projet5	...
1	Pierre	Gascon	4432	AA0232	AA1142				
2	Sophie	Dussault	4302B	AA1142	FG3332	FG4001			
3	Herman	Smith	2212	FG3332	FG4001	FT0023	FG4329	FS7389	
4	Gilles	Dupont	2212	AA0232	FG4329				

**Projets**

Nom	Description
AA0232	Séquençage génome
AA1142	Découverte SNPs
FG3332	Recherche gènes candidats
FG4001	Transgénèse
FG4329	Genome walking
FS7389	Gene silencing
FT0023	Développement banque de cDNA
...	

*Requête SQL ?*

Normalisation: 2. **normalisée****Employés**

#	Prénom	Nom de famille	Local
1	Pierre	Gascon	4432
2	Sophie	Dussault	4302B
3	Herman	Smith	2212
4	Gilles	Dupont	2212

**Projets**

Nom	Description
AA0232	Séquençage génome
AA1142	Découverte SNPs
FG3332	Recherche gènes candidats
FG4001	Transgénèse
FG4329	Genome walking
FS7389	Gene silencing
FT0023	Développement banque de cDNA
...	

1

 $\infty$ **Employés dans projets**

# d'employé	# de projet
1	AA0232
1	AA1142
2	AA1142
2	FG3332
2	FG4001
...	...

 $\infty$ 

1

Normalisation: 1. **non-normalisée**

### Employés\_en\_entreprises

NoEmployé	Prénom	Nom	NoEntreprise	Nom	Adresse	Pays
1	Sven	Torvald	76	DBStore	172 Clearview rd., Randall city	US
2	Pierrette	Fontaine	92	Wal Mart	65 Versant Nord, Québec	Can
3	Rolande	Messier	12	ABC Inc.	1055 Miron, Asbestos	Can
4	Éric	Thomas	92	Wal Mart	65 Versant Nord, Québec	Can
5	Roger	Moore	76	DBStore	172 Clearview rd., Randall city	US
6	Hector	Charland	12	ABC Inc.	1055 Miron, Asbestos	Can

Normalisation: 2. **normalisée****Employés**

NoEmployé	Prénom	Nom	NoEntreprise
1	Sven	Torvald	76
2	Pierrette	Fontaine	92
3	Rolande	Messier	12
4	Éric	Thomas	92
5	Roger	Moore	76
6	Hector	Charland	12

**Entreprises**

NoEntreprise	Nom	Adresse	Pays
12	ABC Inc.	1055 Miron, Asbestos	Can
76	DBStore	172 Clearview rd., Randall city	US
92	Wal Mart	65 Versant Nord, Québec	Can