

L'intron codant du Code de procédure pénale

P^r Jean R. LOBRY

22 mars 2017

Une curiosité scientifique dans le Code de procédure pénale français nous invite à revisiter la notion de codant et de non-codant dans notre ADN. Un brouillon en date du 2017-03-22 en mode RFQ, un acronyme anglais pour « *Request For Comments* »

Table des matières

1	L'intron codant du Législateur	1
1.1	Des Codes et des codes	1
1.1.1	Qu'est-ce qu'un Code juridique?	1
1.1.2	Qu'est-ce qu'un code génétique?	2
1.2	L'article 706-54 du Code de procédure pénale	4
1.2.1	Recherche du segment codant dérogatoire	4
1.2.2	Analyse du segment codant dérogatoire	6
2	Du codant et du non-codant dans les Codes	12
2.1	Bibliographie de l'article 706-54 du Code de procédure pénale . .	12
2.1.1	Textes cités par 706-54	12
2.1.2	Textes citant 706-54	13
2.2	Exploration de la navette parlementaire	13
2.2.1	Codant au lieu de neutre	14
2.2.2	L'exemple des anomalies génétiques	15
2.2.3	Les références aux résolutions du Conseil de l'Union européenne	21

1 L'intron codant du Législateur

1.1 Des Codes et des codes

1.1.1 Qu'est-ce qu'un Code juridique?

POUR un scientifique c'est l'équivalent d'un article prestigieux dit de « *review* » qui fait le point à un instant donné sur l'état de l'art d'un domaine donné. Dans le domaine juridique c'est un « [...] recueil officiel des dispositions législatives et réglementaires qui régissent une matière [...] » [8]. En pratique, c'est un gros volume en papier bible millésimé qui selon que vous préférerez l'édition privée « Dalloz » ou « LexisNexis » sera pour la couverture en rouge, ou bien en bleu, respectivement. Le site de « *legifrance* » offre une version dématérialisée des

Codes mise à jour extemporanément avec un système de suivi de version qui permet même parfois de se projeter dans le futur (*e.g.* sous l'article A38 du Code de procédure pénale on peut dès à présent voir les loci qui seront à renseigner pour l'identification par empreintes génétiques à compter du 1^{er} août 2018). Le site fournit des liens de type référence bibliographiques dans un article scientifique (*e.g.* l'article 706-54 du Code de procédure pénale cite l'article 16-11 du Code civil). On dispose également de l'équivalent du « *Science Citation Index* » (*e.g.* l'article 706-54 du Code de procédure pénale est cité par l'« [a]vis relatif à l'expertise génétique aux fins de détermination des traits morphologiques apparents » de la Commission nationale consultative des droits de l'homme du 17 mars 2016). Les limitations du site sont que l'on ne dispose pas de l'exégèse des juristes des éditions privées et de la correction d'erreurs manifestes (*e.g.* l'article R131-23 du Code pénal qui possède deux « 3^o » dans sa version en vigueur au 1^{er} mars 1994).

NOTEZ que l'on écrira par convention dans ce document un Code juridique avec une majuscule, et ce pour plusieurs raisons :

- 1^o d'un point de vue pragmatique, comme il est question ici de « Codes juridiques » et de « codes génétiques » il est plaisant de pouvoir les distinguer du premier coup d'œil, ne-serait-ce que pour le plaisir d'écrire des titres dans la veine de « Des Codes et des codes » ;
- 2^o un argument d'autorité extrait du site de l'Académie française¹ qui entérine l'usage que « [d]ans tous les titres d'œuvres, le premier terme au moins (ainsi bien sûr que les noms propres) prend la majuscule » et dont une application d'intérêt se trouve dans la section consacrée à l'écriture des nombres en chiffres en l'exemple de « [...] l'article 1246 du Code civil [...] » ;
- 3^o d'un point de vue historique, dans l'édition papier de 1804 du Code civil par Firmin DIDOT on peut voir au niveau du sommaire du TOME PREMIER qu'il y a une majuscule à Code civil (*cf.* figure 5 page 22) ;
- 4^o un argument doctrinal, certes loin d'être issu d'une recherche approfondie, est que les praticiens du droit mettent une majuscule [19, 23, 25, 30, 17, 8, 7, 31, 24, 20, 10, 5, 9, 4, 27, 13, 22, 26].

Il est donc entendu que nous écrirons ici « Code civil », « Code pénal » et « Code de procédure pénale ». Quant à la question de savoir si cela est généralisable à d'autres Codes, il faudrait l'envisager *cum grano salis*. On pourrait en effet être plus réservé pour les Codes dont les parties législatives sont passées sous contrôle gouvernemental², donc de « pseudo-législateurs » pour reprendre l'expression la moins désobligeante³ de François TERRÉ [30].

1.1.2 Qu'est-ce qu'un code génétique ?

SOIENT (1) **C** l'ensemble des $4^3 = 64$ codons possibles dans les séquences codantes et (2) **A** l'ensemble formé de l'union de l'ensemble vide, \emptyset , et des

1. <http://www.academie-francaise.fr/> dernière consultation ... vers la fin de l'éternité !

2. Pour une étude détaillée voir le document de février 2014 intitulé « [l]es ordonnances prises sur le fondement de l'article 38 de la Constitution » disponible sur le site du Sénat.

3. On trouvera dans [30] à propos des codifications dites trompeusement « à droit constant » quelques volées de bois vert bien senties : « frénésie administrative et mandarinale », « programme démentiel de codification », « travail secret de pontifes digne des temps les plus rétrogrades », « démission du Parlement » et « entreprise si vantée par ses auteurs infatués ». Qui a dit que les ouvrages de droit étaient insipides ?

20 acides aminés des protéines, où l'ensemble vide représente les codons non-sens utilisés comme fin de signal de la traduction et les codons non-assignés. Un code génétique est une application surjective de \mathbf{C} dans \mathbf{A} .

$$\mathbf{C} = \{\text{AAA, AAC, AAG, AAT, \dots, TTT}\} \quad (1)$$

$$\mathbf{A} = \{\emptyset, \text{Ala, Arg, \dots, Val}\} \quad (2)$$

La fonction `words()` du paquet  `seqinr` [6] donne si on l'invoque avec la valeur des paramètres par défaut la liste des 64 codons :

```
library(seqinr)
words()
[1] "aaa" "aac" "aag" "aat" "aca" "acc" "acg" "act" "aga" "agc" "agg" "agt" "ata"
[14] "atc" "atg" "att" "caa" "cac" "cag" "cat" "cca" "ccc" "ccg" "cct" "cga" "cgc"
[27] "cgg" "cgt" "cta" "ctc" "ctg" "ctt" "gaa" "gac" "gag" "gat" "gca" "gcc" "gcg"
[40] "gct" "gga" "ggc" "ggg" "ggt" "gta" "gtc" "gtg" "gtt" "taa" "tac" "tag" "tat"
[53] "tca" "tcc" "tcg" "tct" "tga" "tgc" "tgg" "tgt" "tta" "ttc" "ttg" "ttt"
```

La fonction `translate()` permet d'obtenir le code monolettre (*e.g.* K pour la lysine) des acides aminés correspondants. La fonction `aaa()` donne le code en trois lettres (*e.g.* Lys pour la lysine) des acides aminés. La fonction `tablecode()` permet de représenter un code génétique sous la forme d'un tableau tel qu'on le trouve habituellement dans les livres du domaine. Si rien n'est précisé c'est le code génétique « standard » qui est utilisé⁴.

```
sapply(words(), function(x) translate(s2c(x)))
aaa aac aag aat aca acc acg act aga agc agg agt ata atc atg att caa cac cag cat cca
"K" "N" "K" "N" "T" "T" "T" "T" "R" "S" "R" "S" "I" "I" "M" "I" "Q" "H" "Q" "H" "P"
ccc ccg cct cga cgc cgg cgt cta ctc ctg ctt gaa gac gag gat gca gcc gcg gct gga ggc
"p" "p" "p" "R" "R" "R" "R" "L" "L" "L" "L" "E" "D" "E" "D" "A" "A" "A" "A" "G" "G"
ggg ggt gta gtc gtg gtt taa tac tag tat tca tcc tcg tct tga tgc tgg tgt tta ttc ttg
"Q" "G" "V" "V" "V" "V" "*" "Y" "*" "Y" "S" "S" "S" "S" "*" "C" "W" "C" "L" "F" "L"
ttt
"F"

sapply(words(), function(x) aaa(translate(s2c(x))))
aaa aac aag aat aca acc acg act aga agc agg agt ata atc
"Lys" "Asn" "Lys" "Asn" "Thr" "Thr" "Thr" "Thr" "Arg" "Ser" "Arg" "Ser" "Ile" "Ile"
atg att caa cac cag cat cca ccc ccg cct cga cgc cgg cgt
"Met" "Ile" "Gln" "His" "Gln" "His" "Pro" "Pro" "Pro" "Pro" "Arg" "Arg" "Arg" "Arg"
cta ctc ctg ctt gaa gac gag gat gca gcc gcg gct gga ggc
"Leu" "Leu" "Leu" "Leu" "Glu" "Asp" "Glu" "Asp" "Ala" "Ala" "Ala" "Ala" "Gly" "Gly"
ggg ggt gta gtc gtg gtt taa tac tag tat tca tcc tcg tct
"Gly" "Gly" "Val" "Val" "Val" "Val" "Stp" "Tyr" "Stp" "Tyr" "Ser" "Ser" "Ser" "Ser"
tga tgc tgg tgt tta ttc ttg ttt
"Stp" "Cys" "Trp" "Cys" "Leu" "Phe" "Leu" "Phe"
```

```
tablecode()
```

4. Les codes génétiques connus sont disponibles dans ce document : <http://seqinr.r-forge.r-project.org/src/appendix/gencodes.pdf>.

Genetic code 1 : standard							
TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys
TTC	Phe	TCC	Ser	TAC	Tyr	TGC	Cys
TTA	Leu	TCA	Ser	TAA	Stp	TGA	Stp
TTG	Leu	TCG	Ser	TAG	Stp	TGG	Trp
CTT	Leu	CCT	Pro	CAT	His	CGT	Arg
CTC	Leu	CCC	Pro	CAC	His	CGC	Arg
CTA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CTG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser
ATC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
ATA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
ATG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly
GTC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GTA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GTG	Val	GCG	Ala	GAG	Glu	GGG	Gly

La fonction `tablecode()` permet également de produire des tables avec un meilleur rendu typographique en générant du code L^AT_EX. La table 3 page 21 a été produite avec le code  suivant :

```
tablecode(latexfile = "codestd.tex", label = "codestd", size = "small",
caption = "Le code génétique standard")
```

1.2 L'article 706-54 du Code de procédure pénale

1.2.1 Recherche du segment codant dérogoire

LE Code pénal compile à l'attention des « méchants » (*viz.* contrevenants, délinquants, criminels) ce qu'il en coûte que d'enfreindre la loi. Le Code de procédure pénale compile à l'attention des « gentils » (*e.g.* scientifiques, magistrats, policiers) ce qu'ils ont le droit de faire ou de ne pas faire pour neutraliser les « méchants » (*e.g.* le marquage au fer rouge et l'essorillage ne sont plus autorisés pour la prévention de la récidive).

L'ARTICLE 706-54 du Code de procédure pénale dispose dans son avant-dernier alinéa à propos du fichier national automatisé des empreintes génétiques (FNAEG) que les « [...] empreintes génétiques conservées dans ce fichier ne peuvent être réalisées qu'à partir de segments d'acide désoxyribonucléique non codants, à l'exception du segment correspondant au marqueur du sexe. » Il s'agit ici de la version en vigueur au 16 mars 2011, cet alinéa n'existait pas dans la version initiale créée par la loi n° 98-468 du 17 juin 1998, il a été introduit par l'article 29 de la loi n° 2003-239 du 18 mars 2003. Nous avons donc un segment d'ADN marqueur du sexe qui est autorisé à déroger à la règle voulant que seuls les segments non-codants puissent être renseignés dans le FNAEG.

POUR en savoir plus il faut lire le dernier alinéa qui dispose qu'un « [...] décret en Conseil d'État pris après avis de la Commission nationale de l'informatique et des libertés détermine les modalités d'application du présent article. Ce décret précise notamment la durée de conservation des informations enregistrées. » Il nous faut aller maintenant explorer la partie réglementaire du Code de procédure pénale.

C'EST l'article **R53-13** qui dispose que le « [...] nombre et la nature des segments d'ADN non codants sur lesquels portent les analyses d'identification par empreintes génétiques sont définis par arrêté du ministre de la justice et du ministre de l'intérieur pris après avis de la commission chargée d'agréer les personnes habilitées à effectuer des missions d'identification par empreintes génétiques dans le cadre des procédures judiciaires, prévue par l'article 1^{er} du décret n° 97-109 du 6 février 1997 relatif aux conditions d'agrément des personnes habilitées à procéder à des identifications par empreintes génétiques dans le cadre d'une procédure judiciaire. » Il nous faut donc explorer les arrêtés.

C'EST finalement l'article **A38** qui dispose dans la version en vigueur avec terme du 22 août 2015 au 1^{er} août 2018 que les « [...] analyses destinées à l'identification génétique peuvent ne porter que sur les seuls segments d'ADN figurant dans le tableau ci-après :

Segment d'ADN ou « locus » selon la nomenclature internationale	Localisation
D3S1358	Chromosome 3
VWA	Chromosome 12
D8S1179	Chromosome 8
D21S11	Chromosome 21
D18S51	Chromosome 18
TH01	Chromosome 11
FGA	Chromosome 4
D16S539	Chromosome 16
D5S818	Chromosome 5
D13S317	Chromosome 13
D7S820	Chromosome 7
CSF1PO	Chromosome 5
TPOX	Chromosome 2
Amélogénine	Chromosomes X et Y

Elles portent alors également sur deux au moins des quatre segments d'ADN suivants :

Segment d'ADN ou « locus » selon la nomenclature internationale	Localisation
D2S1338	Chromosome 2
D19S433	Chromosome 19
Penta E	Chromosome 15
Penta D	Chromosome 21

Elles peuvent également porter sur le segment suivant :

Segment d'ADN ou « locus » selon la nomenclature internationale	Localisation
SE33 (ACTBP2)	Chromosome 6

Le segment d'ADN codant dérogatoire marqueur du sexe, selon l'avant-dernier alinéa de l'article 706-54 du Code de procédure pénale, est donc l'amélogénine puisque c'est le seul locus porté par les chromosomes sexuels X et Y.

1.2.2 Analyse du segment codant dérogatoire

L'AMÉLOGÉNINE est une famille de protéines impliquées dans le développement de l'émail des dents. Le gène codant pour l'amélogénine est localisé sur le chromosome X (AMELX) et sur le chromosome Y (AMELY) [21]. L'amélogénine existe sous plusieurs isoformes en raison de la possibilité d'épissages alternatifs des transcrits du gène. Le gène n'est pas situé sur les régions pseudo-autosomiques des chromosomes sexuels, il n'y a donc pas de possibilité de recombinaison entre AMELX et AMELY. Cette propriété est très intéressante puisqu'elle signifie que ces deux allèles évoluent indépendamment dans les lignées mâles et femelles.

Le premier intron interne d'AMELX présente une délétion de 6 paires de bases par rapport à celui d'AMELY. Récupérons les séquences et les annotations de ces allèles et sauvegardons les résultats pour un usage ultérieur :

```
choosebank("genbank")
amelx.list <- query("amelx", "N=HUMAMELX")
amely.list <- query("amely", "N=HUMAMELY")
amel <- list() ; amel$X <- list() ; amel$Y <- list()
amel$X$ann <- getAnnot(amelx.list)[[1]]
amel$X$seq <- getSequence(amelx.list)[[1]]
amel$Y$ann <- getAnnot(amely.list)[[1]]
amel$Y$seq <- getSequence(amely.list)[[1]]
save(amel, file = "amel.rda")
```

L'EXAMEN des annotations d'AMELX montre que le premier intron interne se situe entre les positions 240 et 1648 :

```
load("amel.rda")
cat(amel$X$ann, sep = "\n")
LOCUS           HUMAMELX                2872 bp    DNA     linear   PRI 06-MAR-1995
DEFINITION      Human amelogenin (AMELX) gene, 3' end of cds.
ACCESSION       M55418 J04777
VERSION         M55418.1 GI:178526
KEYWORDS        AMELX gene; amelogenin; major extracellular matrix protein.
SOURCE          Homo sapiens (human)
ORGANISM        Homo sapiens
                 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                 Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
                 Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 2872)
AUTHORS         Nakahori,Y., Takenaka,O. and Nakagome,Y.
TITLE           A human X-Y homologous region encodes 'amelogenin'
JOURNAL         Genomics 9 (2), 264-269 (1991)
PUBMED          2004775
COMMENT         Original source text: Human lymphocyte cell line GM1416 DNA, clone
                 pYN88001.
FEATURES        Location/Qualifiers
                 source                1..2872
                 /organism="Homo sapiens"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:9606"
                 /map="Xp22.31-p22.1"
                 /clone="pYN88001"
                 /cell_line="GM1416"
                 /cell_type="lymphocyte"
                 /tissue_lib="Human mutant cell repository, New Jersey"
                 intron                1..191
                 /gene="AMELX"
                 /note="G00-119-675"
                 /number=1
                 gene                  join(192..239,1649..1690,1961..2401)
                 /gene="AMELX"
                 CDS                    join(192..239,1649..1690,1961..2401)
```

```

/ gene="AMELX"
/ codon_start=1
/ product="amelogenin"
/ protein_id="AAA62826.1"
/ db_xref="GI:178527"
/ db_xref="GDB:G00-119-675"
/ translation="LPPHPGHPGYINFSYEVLTPLKWKYSIRPPYPSYGYEPMGGWLH
HQIIPVLSQQHPPTHTLQPHHHIPVVPAQQPVIPQQPMPVPGQHSMTPIQHHQPNLP
PPAQQPYQPQPVPQPHQPMQPPVHPMQPLPPQPPLPMPFMPQLPPLMLDLTLEA
WPSTDKTKREEVVSIF"
exon      192..239
/ gene="AMELX"
/ note="G00-119-675"
/ number=2
intron    240..1648
/ gene="AMELX"
/ note="G00-119-675"
/ number=2
exon      1649..1690
/ gene="AMELX"
/ note="G00-119-675"
/ number=3
intron    1691..1960
/ gene="AMELX"
/ note="G00-119-675"
/ number=3
exon      1961..>2401
/ gene="AMELX"
/ note="G00-119-675"
ORIGIN    Chromosome Xp22.31-p22.1.

```

L'EXAMEN des annotations d'AMELY montre que le premier intron interne se situe entre les positions 240 et 1466 :

```

cat(amel$Y$ann, sep = "\n")
LOCUS      HUMAMELY                      3272 bp    DNA     linear   PRI 06-MAR-1995
DEFINITION Human amelogenin (AMELY) gene, 3' end of cds.
ACCESSION  M55419 J04777 J04778
VERSION    M55419.1  GI:178524
KEYWORDS   AMELY gene; amelogenin; major extracellular matrix protein.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 3272)
AUTHORS    Nakahori,Y., Takenaka,O. and Nakagome,Y.
TITLE      A human X-Y homologous region encodes 'amelogenin'
JOURNAL    Genomics 9 (2), 264-269 (1991)
PUBMED    2004775
COMMENT    Original source text: Human (ATCC 57735) DNA, clone 87-4.
FEATURES   Location/Qualifiers
            source                1..3272
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /map="Yp11.2"
            /clone="87-4"
            /tissue_lib="ATCC 57735"
            intron                <1..191
            /gene="AMELY"
            /note="G00-119-676"
            /number=1
            gene                  join(192..239,1467..1511,1781..2221)
            /gene="AMELY"
            CDS                   join(192..239,1467..1511,1781..2221)
            /gene="AMELY"
            /codon_start=1
            /product="amelogenin"
            /protein_id="AAA62827.1"
            /db_xref="GI:178525"
            /db_xref="GDB:G00-119-676"
            /translation="LPPHPGHPGYINFSYEVLTPLKWKYSMIRPPYSSYGYEPMGGWL
            HHQIIPVVSQQHPLTHTLQSHHHIPVVPAQQPRVRQALMPVPGQSQSMTPTQHHQPNL
            PLPAQQPFQPPVQPHQPMQPPVHPMQPLLPQPPLPMPFPLRPLPPLDLHLHE
            AWPATDKTKQE EVVSTP"
            exon                  192..239

```

```

                /gene="AMELY"
                /note="G00-119-676"
                /number=2
intron         240..1466
                /gene="AMELY"
                /note="G00-119-676"
                /number=2
exon           1467..1511
                /gene="AMELY"
                /note="G00-119-676"
                /number=3
intron         1512..1780
                /gene="AMELY"
                /note="G00-119-676"
                /number=3
exon           1781..>2221
                /gene="AMELY"
                /note="G00-119-676"
ORIGIN         Chromosome Y.

```

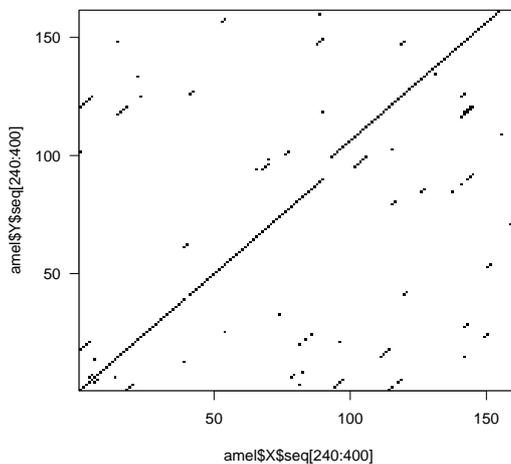
La représentation graphique de type « *dot plot* » [16] permet de mettre facilement en évidence la délétion dans l'allèle AMELX. Dans sa forme la plus simple un point est représenté à la position (i, j) si, et seulement si, le i^{e} caractère de la première séquence est le même que le j^{e} caractère de la seconde séquence. Des formes plus élaborées utilisent une fenêtre glissante et une valeur seuil contrôlant la représentation d'un couple de fenêtres. Deux séquences identiques donnent une droite d'équation $y = x$, les insertions et délétions se caractérisent par des saut verticaux ou horizontaux.

```

dotPlot(amel$X$seq[240:400], amel$Y$seq[240:400],10,1,7, las = 1,
        main = "La délétion au début du premier intron interne de AMELX")

```

La délétion au début du premier intron interne de AMELX



C'EST cette délétion qui est la cible des « kits » de biologie moléculaire pour la détermination du sexe en sciences forensiques. Ces « kits » utilisent comme amorces celles de Sullivan *et al.* [29] ou bien des variants mineurs de ces dernières, *cf.* sur le site « STRBase »⁵ de John M. BUTLER la page dédiée à l'amélogénine. On trouve sur ce même site le détail des amorces utilisées :

5. <http://www.cstl.nist.gov/strbase/index.htm>

STR Locus	Chromosomal Position	GenBank Accession	Primer Sequence (5'→3')	Reference
Amelogenin	Xp22.1-.3 Yp11.2	M55418 M55419	5'-CCCTGGGCTCTGTAAAGAATAGTG-3' 5'-ATCAGAGCTTAAACTGGGAAGCTG-3'	[29, 18]

LA colonne « GenBank Accession » est une clef d'identification, on vérifie ici que M55418 et M55419 correspondent bien aux séquences que nous avons récupérées précédemment. Essayons de localiser ces amorces dans nos séquences, on commence par les définir dans les objets `amo1` et `amo2` :

```
(amo1 <- tolower("CCCTGGGCTCTGTAAAGAATAGTG"))
[1] "ccctgggctctgtaaagaatagtg"
(amo2 <- tolower("ATCAGAGCTTAAACTGGGAAGCTG"))
[1] "atcagagcttaaaactgggaagctg"
```

MAIS attention ici : les amorces sont données conventionnellement dans le sens 5'→3', nous avons donc besoin de travailler avec le brin complémentaire pour la deuxième amorce :

```
cplt <- function(x) c2s(comp(rev(s2c(x))))
(amo2 <- cplt(amo2))
[1] "cagcttcccagtttaagctctgat"
```

ON peut maintenant rechercher la position des amorces dans les séquences pour en déduire la taille des fragments amplifiés :

```
position <- function(amo, adn) gregexpr(amo, c2s(adn))[[1]][1]
(amel$X$amo1 <- position(amo1, amel$X$seq))
[1] 287
(amel$X$amo2 <- position(amo2, amel$X$seq))
[1] 369
(amel$Y$amo1 <- position(amo1, amel$Y$seq))
[1] 287
(amel$Y$amo2 <- position(amo2, amel$Y$seq))
[1] 375
(ampliX <- amel$X$amo2 - amel$X$amo1 + nchar(amo2))
[1] 106
(ampliY <- amel$Y$amo2 - amel$Y$amo1 + nchar(amo2))
[1] 112
```

ON aura donc un amplicon de 106 paires de bases sur le chromosome X et de 112 paires de bases sur le chromosome Y, cela correspond bien à la délétion de 6 paires de bases sur le chromosome X par rapport au chromosome Y. Comme les femelles possèdent deux chromosomes X on observera un seul pic pour un échantillon de leur ADN. Les mâles ayant un chromosome X et un chromosome Y on s'attend à avoir deux pics pour un échantillon de leur ADN. Voyons un exemple concret avec les résultats obtenus avec un « kit [2] » dont l'échelle allélique est donné dans la figure 1 page 10. L'échantillon d'ADN analysé

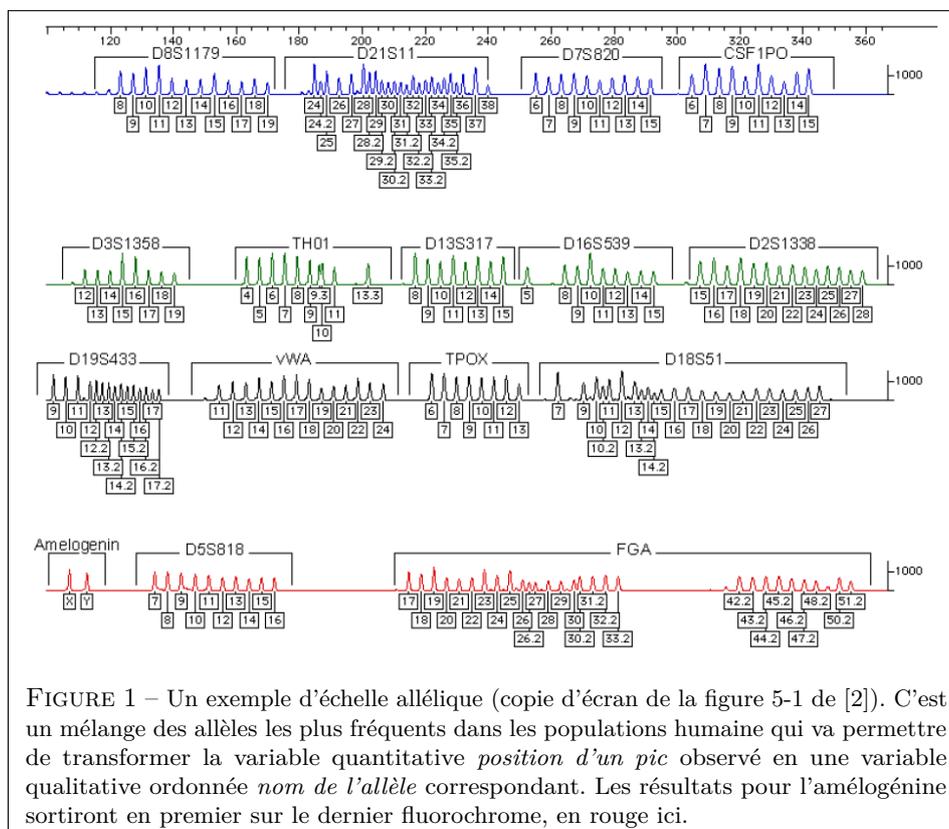


FIGURE 1 – Un exemple d'échelle allélique (copie d'écran de la figure 5-1 de [2]). C'est un mélange des allèles les plus fréquents dans les populations humaine qui va permettre de transformer la variable quantitative *position d'un pic* observé en une variable qualitative ordonnée *nom de l'allèle* correspondant. Les résultats pour l'amélogénine sortiront en premier sur le dernier fluorochrome, en rouge ici.

est celui de l'auteur de ce document⁶. Les résultats pour le dernier fluorochrome sont dans DATA.4 :

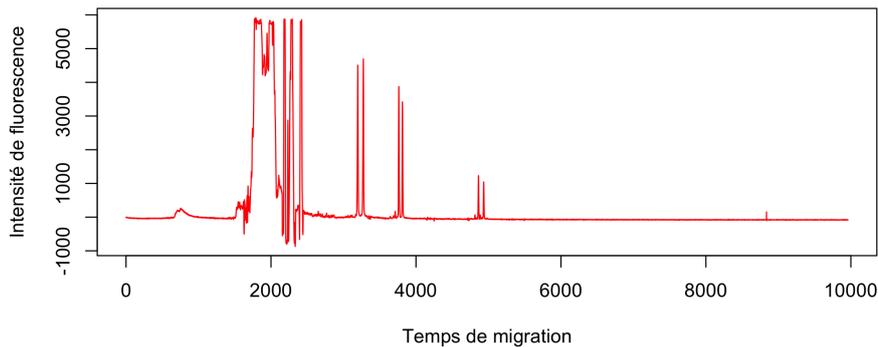
```
data(JLO)
names(JLO$Data)
[1] "ANME.1" "CTID.1" "CTNM.1" "CTOw.1" "CTTL.1" "CpEP.1" "DATA.1"
[8] "DATA.2" "DATA.3" "DATA.4" "DATA.5" "DATA.6" "DATA.7" "DATA.8"
[15] "DATA.105" "DCHT.1" "DSam.1" "DySN.1" "Dye#.1" "DyeB.1" "DyeB.2"
[22] "DyeB.3" "DyeB.4" "DyeB.5" "DyeN.1" "DyeN.2" "DyeN.3" "DyeN.4"
[29] "DyeN.5" "DyeW.1" "DyeW.2" "DyeW.3" "DyeW.4" "DyeW.5" "EPVt.1"
[36] "EVNT.1" "EVNT.2" "EVNT.3" "EVNT.4" "GTyp.1" "HCFG.1" "HCFG.2"
[43] "HCFG.3" "HCFG.4" "InSc.1" "InVt.1" "LANE.1" "LIMS.1" "LNTD.1"
[50] "LsrP.1" "MCHN.1" "MODF.1" "MODL.1" "NAV.1" "NLNE.1" "OfSc.1"
[57] "PANL.1" "PSZE.1" "PTYP.1" "PXLB.1" "RGNm.1" "RMXV.1" "RMDN.1"
[64] "RMDV.1" "RMDX.1" "RPrN.1" "RPrV.1" "RUND.1" "RUND.2" "RUND.3"
[71] "RUND.4" "RUNT.1" "RUNT.2" "RUNT.3" "RUNT.4" "Rate.1" "RunN.1"
```

6. Les données complètes sont distribuées dans le paquet `seqinr` [6]. Les paquets  sont disponibles sur les nombreux miroirs du CRAN (acronyme pour « *the Comprehensive R Archiv Network* ») : 1 en Algérie, 1 en Argentine, 3 en Australie, 2 en Belgique, 3 au Brésil, 1 en Bulgarie, 4 au Canada, 2 au Chili, 4 en Chine, 1 en Colombie, 1 en Tchèque, 1 au Danemark, 1 en Équateur, 1 en El Salvador, 1 en Estonie, 6 en France, 2 en Allemagne, 1 en Grèce, 1 en Hongrie, 1 en Islande, 1 en Inde, 1 en Indonésie, 1 en Iran, 1 en Irlande, 3 en Italie, 1 en Japon, 3 en Corée, 1 au Liban, 1 en Malaisie, 2 au Mexique, 2 aux Pays-bas, 1 en Nouvelle-Zélande, 1 en Norvège, 1 aux Philippines, 2 au Portugal, 1 en Russie, 1 en Serbie, 1 à Singapour, 2 en Afrique du Sud, 2 en Espagne, 1 en Suède, 1 en Suisse, 2 à Taiwan, 1 en Thaïlande, 1 en Turquie, 5 au Royaume-Uni, 16 aux États Unis d'Amérique et 1 au Vénézuéla. Il n'est donc pas déraisonnable de penser que l'empreinte génétique de l'auteur de ce document est la plus fichée au monde. Dernière consultation du CRAN le 2017-03-06 avec neutralisation des doubles comptes entre les protocoles `http` et `https`.

```
[78] "SCAN.1" "SMED.1" "SMLt.1" "STYP.1" "SVER.1" "SVER.3" "SVER.4"
[85] "Satd.1" "Scal.1" "Scan.1" "SpNm.1" "StdF.1" "TUBE.1" "Tmpr.1"
[92] "UDEF.1" "User.1"
```

```
plot(JLO$Data$DATA.4, type = "l", xlab = "Temps de migration",
      ylab = "Intensité de fluorescence", col = "red",
      main = "Résultat de l'électrophorèse pour le dernier fluorochrome")
```

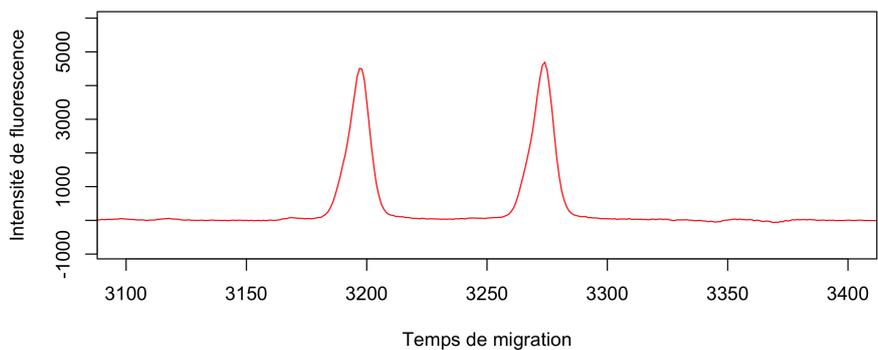
Résultat de l'électrophorèse pour le dernier fluorochrome



ZOOMONS sur la région utile pour bien mettre en évidence les résultats pour l'amélogénine :

```
plot(JLO$Data$DATA.4, type = "l", xlab = "Temps de migration",
      ylab = "Intensité de fluorescence", col = "red",
      main = "Résultat de l'électrophorèse pour le dernier fluorochrome",
      xlim = c(3100, 3400))
```

Résultat de l'électrophorèse pour le dernier fluorochrome



POUR le locus de l'amélogénine nous avons deux pics, ce qui signifie que l'échantillon d'ADN analysé provient d'un mâle. La lecture des résultats pour l'amélogénine est donc particulièrement simple, il n'en serait pas de même pour les autres loci ou une étape de calibration est nécessaire en amont pour pouvoir identifier les allèles⁷.

7. Pour plus de détails voir le document « Lecture de fichiers au format ABIF sous R » à <https://pbil.univ-lyon1.fr/R/pdf/for10.pdf>

Source	Référence
Code civil	art. 16-11
Code de procédure pénale	art. 74
Code de procédure pénale	art. 80-4
Code de procédure pénale	art. 706-55
Code de procédure pénale	art. 706-120
Code de procédure pénale	art. 706-125
Code de procédure pénale	art. 706-129
Code de procédure pénale	art. 706-133
Code de procédure pénale	art. 706-134

TABLE 1 – Les textes cités par l'article 706-54 du Code de procédure pénal d'après le site de [legifrance](http://legifrance.gouv.fr) consulté le 2017-03-10.

La localisation du fragment amplifié utilisé pour la détermination du sexe est intégralement situé sur le premier intron du gène : pour AMELX de 287 à 369 (intron de 240 à 1648) . et pour AMELY de 287 à 374 (intron de 240 à 1466). Or, par définition, les introns sont éliminés lors de la phase d'épissage, ils ne sont jamais traduits, **ce ne sont pas des séquences codantes**. Comment en est-on arrivé à ce que le Législateur affirme le rebours dans l'article 706-54 du Code de procédure pénale ? Est-ce une confusion entre transcription et traduction, entre génotype et phénotype, entre séquence neutre et séquence soumise à la sélection naturelle ? Seule une exploration des textes et débats des assemblées est susceptible de nous éclairer.

On peut sans doute éliminer l'hypothèse d'une confusion entre la transcription et la traduction puisque parmi les loci autorisés par l'article A38 du Code de procédure pénale on en trouve plusieurs qui sont transcrits (*viz.* VWA est situé dans le 40^e intron du gène codant pour le facteur de von WILLEBRAND, TH01 sur le premier intron du gène de la tyrosine hydroxylase, FGA sur le troisième intron du gène pour la chaîne α du fibrinogène, CSF1PO sur le 6^e intron du gène pour le proto-oncogene c-fms, TPOX sur le 10^e intron du gène pour la peroxydase thyroïdienne).

2 Du codant et du non-codant dans les Codes

2.1 Bibliographie de l'article 706-54 du Code de procédure pénale

2.1.1 Textes cités par 706-54

L'ARTICLE 706-54 fait référence à 9 textes donnés dans la table 1 page 12. Aucun de ces articles ne parle de séquences codantes ou non codantes : ils précisent les conditions nécessaires pour figurer au FNAEG. Huit sont des références internes au Code de procédure pénale et un est une référence externe à l'article 16-11 du Code civil qui dispose des cas où l'identification d'une personne par ses empreintes génétiques est possible.

Source	Référence
Conseil constitutionnel	Décision QPC n° 2010-25
CNIL	Délibération n° 2008-113
CNIL	Délibération n° 2012-408
Code de procédure pénale	art. 706-56
Code de procédure pénale	art. R53-9
Code de procédure pénale	art. R53-13-1
Décret	n° 2009-785
Arrêté	10 août 2015
CNCDH	Avis du 17 mars 2016

TABLE 2 – Les textes citant l'article 706-54 du Code de procédure pénal d'après le site de **legifrance** consulté le 2017-03-10. Dans la colonne « Source » CNIL et CNCDH sont des acronymes pour « Commission Nationale de l'Informatique et des Libertés » et « Commission Nationale Consultative des Droits de l'Homme », respectivement

2.1.2 Textes citant 706-54

Ils sont donnés dans la table 2 page 13

2.2 Exploration de la navette parlementaire

L'AVANT-DERNIER alinéa de l'article 706-54 du Code de procédure pénale a été introduit par l'article 29 de la loi n° 2003-239 du 18 mars 2003 pour la sécurité intérieure⁸. Le tableau de concordance entre les articles de la loi et les articles examinés en cours de discussion donné sur le site de l'Assemblée nationale nous indique que nous devons nous intéresser à l'article 15 du projet⁹ de loi.

DANS l'exposé des motifs du projet de loi déposé au Sénat le 23 octobre 2002 par M. Nicolas SARKOZY, ministre de l'intérieur, de la sécurité intérieure et des libertés locales on trouve l'explication suivante : « Il est par ailleurs important de préciser dans l'article 706-54 que seuls les segments non codants de l'ADN, à l'exception de celui correspondant au marqueur du sexe, sont utilisés pour le fonctionnement du FNAEG. Cette précision, essentielle, qui ne figure actuellement que dans une disposition réglementaire (article R. 53-13 du code de procédure pénale), constitue une garantie forte pour les libertés publiques puisqu'aucune caractéristique physique des personnes inscrites au fichier, à part le sexe, ne figurera dans le fichier. »

VOILÀ qui est fort instructif. L'objectif, fort louable au demeurant, est donc de protéger les libertés publiques en excluant tout trait phénotypique, c'est à dire toute « caractéristique physique des personnes » du FNAEG. On peut comprendre l'erreur de raisonnement ainsi : le sexe des personnes est à l'évidence un trait phénotypique, donc un segment d'ADN qui prédit le sexe devrait être forcément codant. C'est archi-faux d'un point de vue scientifique : ce n'est pas parce qu'un locus est non-codant qu'il est sans effet sur le phénotype. Par

⁸. Publiée au journal officiel du 19 mars 2003.

⁹. Les projets de loi sont d'origine gouvernementale, les propositions de loi d'origine parlementaire.

exemple, une mutation dans la séquence promotrice de la transcription d'un gène, qui n'est ni transcrite ni traduite, peut avoir des effets majeurs sur le phénotype.

DANS le rapport déposé au nom de la commission des lois par M. Christian ESTROSI le 18 décembre 2002 on retrouve la même justification : « [i] est précisé que les empreintes conservées dans le fichier ne pourront être réalisées qu'à partir de segments d'ADN non codants, à l'exception du segment correspondant au marqueur du sexe. Jusqu'à présent, cette exigence, sur laquelle le rapporteur reviendra, n'avait été fixée que par la voie réglementaire (article R. 53-13 du code de procédure pénale) : son importance justifiait qu'elle soit consacrée par le législateur. »

LE rapporteur précise plus loin : « [c]omme on l'a vu, il est également explicitement prévu que l'ADN analysé ne sera pas codant : il ne s'agit donc que d'un code sans signification biologique, les analyses ne pouvant s'effectuer sur les segments spécifiques permettant, par exemple, de déterminer l'existence d'anomalies génétiques. Cette précaution est d'ailleurs requise par la recommandation R(92)1 du comité des ministres du Conseil de l'Europe et les résolutions du 9 juin 1997 et du 25 juin 2001 du Conseil de l'Union européenne. Il n'est pas inutile, à ce stade, de rappeler les termes de cette dernière résolution, auxquels la France se conforme scrupuleusement : "Lors de l'échange des résultats des analyses d'ADN, les États membres sont instamment invités à limiter les résultats des analyses d'ADN aux segments chromosomiques ne contenant aucun facteur d'expression de l'information génétique, c'est-à-dire ne fournissant pas, en l'état actuel des connaissances, d'informations sur des caractéristiques héréditaires spécifiques. (...) Si l'évolution scientifique venait à révéler que l'un des marqueurs d'ADN recommandés dans la présente résolution fournit des informations sur des caractéristiques héréditaires spécifiques, il serait recommandé aux États membres de ne plus utiliser ce marqueur lors de l'échange des résultats des analyses d'ADN". » Les précisions du rapporteur sont très éclairantes sur les trois points développés ci-après.

2.2.1 Codant au lieu de neutre

L'EXPRESSION « un code sans signification biologique » est typique de la confusion que l'on rencontre souvent chez les journalistes entre génome et code génétique. Par exemple¹⁰ dans un article publié le 8 mars 2017 à 18h23 par Pierre HEDRICH sur le site du *Nouvel Observateur* on lit : « [...] territoire des bonobos, cette espèce de grands singes dont l'homme partage 98,7 % du code génétique. » Ce n'est pas correct : nous partageons 98,7 % de notre génome avec les bonobos, mais pour les codes génétiques nous utilisons à 100 % les mêmes (*cf.* figure 2 page 15). Les deux termes ne sont pas synonymes : un génome est le support de l'information génétique alors qu'un code génétique est une application surjective (*cf.* page 2). Il semble donc que l'on ait ici la même confusion ici et que « non codant » signifie en fait neutre d'un point de vue évolutif : on ne s'autorise à renseigner le FNAEG qu'avec des régions du génome qui ne sont pas soumises à une pression de sélection. Les régions neutres ne servent à rien d'un point de vue évolutif dans le sens où leurs mutations ne modulent ni en bien ni en mal le succès reproductif de ceux qui les portent.

10. Exemple pris au hasard, on en trouvera de nombreux autres en cherchant « code génétique » dans un moteur de recherche spécialisé dans l'indexation des journaux en ligne.

La bonne nouvelle ici c'est que les régions neutres sont hypervariables, donc particulièrement bien appropriées à l'objectif d'identification des individus. La mauvaise nouvelle c'est que ce n'est pas parce qu'une région est neutre qu'elle ne peut pas nous renseigner, indirectement, sur un trait phénotypique. C'est cette trivialité de la génétique que semble découvrir avec un brin d'effroi le garde des Sceaux dans une dépêche¹¹ du 29 juin 2011 : « [d]'autre part, même si elles sont mises en œuvre à partir de segments d'acide désoxyribonucléique [sic] (ADN) non codants conformément aux exigences de l'avant dernier alinéa de l'article 706-54 du code de procédure pénale, ces méthodes sont indirectement susceptibles de renseigner sur les caractères apparents d'une personne et constituer ainsi un examen de ses caractéristiques génétiques. »

TTT Phe	TCT Ser	TAT Tyr	TGT Cys	TTT Phe	TCT Ser	TAT Tyr	TGT Cys
TTC Phe	TCC Ser	TAC Tyr	TGC Cys	TTC Phe	TCC Ser	TAC Tyr	TGC Cys
TTA Leu	TCA Ser	TAA Stp	TGA Stp	TTA Leu	TCA Ser	TAA Stp	TGA Trp
TTG Leu	TCG Ser	TAG Stp	TGG Trp	TTG Leu	TCG Ser	TAG Stp	TGG Trp
CTT Leu	CCT Pro	CAT His	CGT Arg	CTT Leu	CCT Pro	CAT His	CGT Arg
CTC Leu	CCC Pro	CAC His	CGC Arg	CTC Leu	CCC Pro	CAC His	CGC Arg
CTA Leu	CCA Pro	CAA Gln	CGA Arg	CTA Leu	CCA Pro	CAA Gln	CGA Arg
CTG Leu	CCG Pro	CAG Gln	CGG Arg	CTG Leu	CCG Pro	CAG Gln	CGG Arg
ATT Ile	ACT Thr	AAT Asn	AGT Ser	ATT Ile	ACT Thr	AAT Asn	AGT Ser
ATC Ile	ACC Thr	AAC Asn	AGC Ser	ATC Ile	ACC Thr	AAC Asn	AGC Ser
ATA Ile	ACA Thr	AAA Lys	AGA Arg	ATA Met	ACA Thr	AAA Lys	AGA Stp
ATG Met	ACG Thr	AAG Lys	AGG Arg	ATG Met	ACG Thr	AAG Lys	AGG Stp
GTT Val	GCT Ala	GAT Asp	GGT Gly	GTT Val	GCT Ala	GAT Asp	GGT Gly
GTC Val	GCC Ala	GAC Asp	GGC Gly	GTC Val	GCC Ala	GAC Asp	GGC Gly
GTA Val	GCA Ala	GAA Glu	GGA Gly	GTA Val	GCA Ala	GAA Glu	GGA Gly
GTG Val	GCG Ala	GAG Glu	GGG Gly	GTG Val	GCG Ala	GAG Glu	GGG Gly

FIGURE 2 – Les deux codes génétiques utilisés par l'homme et le bonobo sont tritement identiques. À gauche le code utilisé pour traduire les gènes à localisation nucléaire (*i.e.* qui sont dans le compartiment sub-cellulaire appelé le noyau) et à droite ceux à localisation mitochondriale (*i.e.* qui sont dans le compartiment sub-cellulaire appelé la mitochondrie). Les différences entre ces deux codes sont indiquées en rouge. On peut généraliser l'assertion en disant qu'en l'état actuel des connaissances tous les animaux vertébrés utilisent à 100 % ces deux codes génétiques. Par conséquent, en tant que membre de l'espèce *Homo sapiens*, nos codes génétiques sont à 100 % identiques à ceux du poisson rouge « BUBULLE », de la souris « CHEESE », du hérisson « KISSIFROTT », de la mouette « rieuse » et du chat « dingue » de Gaston LAGAFFE. Les codes génétiques utilisés par le marsupilami ne sont pas documentés à ce jour : <http://seqinr.r-forge.r-project.org/src/appendix/gencodes.pdf>, connaissant l'animal, il se pourrait qu'une exception soit à venir.

2.2.2 L'exemple des anomalies génétiques

Le rapporteur précise que « les analyses ne [peuvent] s'effectuer sur les segments spécifiques permettant, par exemple, de déterminer l'existence d'anomalies génétiques. » Ce n'est qu'un exemple mais il est d'importance.

PRENONS l'exemple du locus TH01 dont l'utilisation pour l'identification par empreintes génétiques est autorisée par l'article A38 du Code de procédure

11. Il s'agit de la dépêche référencée « CRIM-PJ N° 08-28.H5 tome 4 »

pénale. Ce locus, non codant, est situé sur le premier intron du gène codant pour la tyrosine hydroxylase. Cet enzyme catalysant la réaction limitante pour la biosynthèse de la catécholamine, son gène est un bon candidat pour l'étiologie de l'hypertension artérielle. Pankaj SHARMA *et al.* [28] ont cherché si certains allèles du locus TH01 ne pourraient pas être associés avec l'hypertension essentielle¹². La lecture de [28] est difficile parce que les auteurs n'utilisent pas la nomenclature internationale [3] pour désigner les allèles observés mais les lettres A, B, C, D et E. Dans la partie matériel et méthodes la taille des amplicons est donné comme variant de 313 à 329 bp mais dans la partie résultats les allèles observés sont donnés à 311, 315, 319, 323 et 327 bp, respectivement. Il ont tous la même taille modulo 4 ce qui est cohérent avec un microsatellite à base de tétranucléotides :

```
lal <- c(311, 315, 319, 323, 327)
lal %% 4
[1] 3 3 3 3 3
```

RÉCUPÉrons la séquence et les annotations du gène codant pour la tyrosine hydroxylase et sauvegardons les résultats :

```
library(seqinr)
choosebank("genbank")
th01.list <- query("th01", "AC=D00269")
th01 <- list()
th01$ann <- getAnnot(th01.list, nbl = 500)[[1]]
th01$seq <- getSequence(th01.list)[[1]]
save(th01, file = "th01.rda")
```

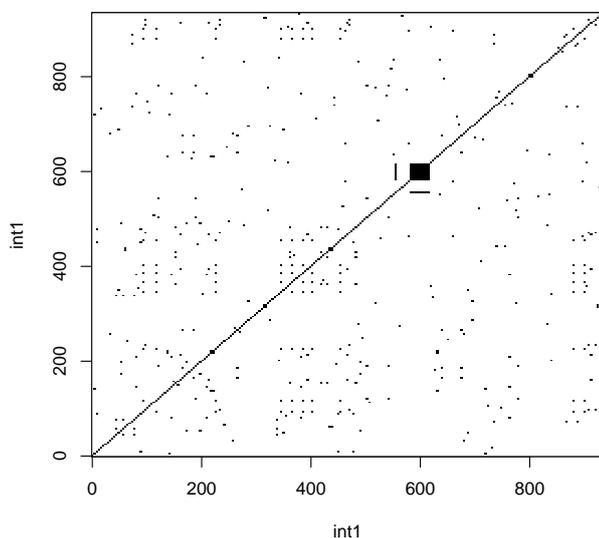
LA lecture des annotations nous permet d'extraire la séquence du premier intron puis en le confrontant avec lui-même dans une représentation de type « *dotplot* » de repérer immédiatement la localisation des segments répétés. L'examen de la séquence permet de mettre en évidence 9 répétitions du tétranucléotide TCAT, l'allèle de la séquence de genbank est donc dans la nomenclature internationale le 9.

```
load("th01.rda")
cat(th01$ann[c(1:3, 105:106)], sep = "\n")
LOCUS      D00269                7854 bp    DNA     linear   PRI 26-JUL-2016
DEFINITION Homo sapiens gene for tyrosine hydroxylase, partial cds,
            alternative splicing forms.
            intron           591..1523
                               /number=1

int1 <- th01$seq[591:1523]
dotPlot(int1, int1, 4, 4, 4)
splitseq(int1[552:640], word = 4)

[1] "ccat" "tggc" "ctgt" "tcct" "ccct" "tatt" "tccc" "tcat" "tcat" "tcat" "tcat"
[12] "tcat" "tcat" "tcat" "tcat" "tcat" "tcac" "catg" "gagt" "ctgt" "gttc" "cctg"
```

12. Un terme savant pour dire que dans la majorité des cas (plus de 95 %) on ne connaît pas de cause précise expliquant l'élévation de la pression artérielle ; « essentielle » signifie ici « de la plus grande importance ».



CALCULONS la taille des amplicons avec les amorces utilisées par Pankaj SHARMA *et al.* [28].

```
(Sharma1 <- tolower("TCCAAAAATCCAAGATGGC"))
[1] "tccaaaaatccaagatggc"
(Sharma2 <- tolower("ACAGGGAACACAGACTCCATG"))
[1] "acaggaacacagactccatg"
(Sharma2 <- cplt(Sharma2))
[1] "catggagtctgtgtccctgt"
position(Sharma2, int1) - position(Sharma1, int1) + nchar(Sharma2)
[1] 323
```

ON en déduit que l'allèle D correspond à l'allèle 9. Les allèles A, B, C, D et E de Pankaj SHARMA *et al.* [28] correspondent donc aux allèles 6, 7, 8, 9 et 10 dans la nomenclature internationale [3]. Je trouve curieux que les auteurs n'aient observé aucune occurrence de l'allèle 9.3 (326 bp) sur 433 individus étudiés : c'est un allèle fréquent¹³ avec une fréquence médiane de 20 % chez 137 populations humaines et un écart inter-quartiles de 20 %. Même dans une population du Mozambique où l'allèle 9.3 est très rare [1] on en trouve une proportion de 2.3 %.

Les résultats bruts sont donnés dans la figure 3 page 18. Importer les données sous R :

```
dta <- read.table("SharmaP.txt", header = TRUE, sep = "\t", row.names = 1)
dta
  AA AB AC AD AE BB BC BD BE CC CD CE DD DE EE Total
HT 10 22 13 15 40 14 7 6 22 3 7 18 2 16 32 227
NT 11 12 18 26 23 3 6 18 27 4 9 11 8 14 16 206
chisq.test(dta[, 1:15], simulate = TRUE)
```

13. Pour une représentation graphique des données compilées dans la base `OmniPop200.1.xls` par Brian BURRITT du département de police à San Diego (CA, USA) voir la fiche « [l]es fréquences alléliques dans les populations humaines des microsatellites utilisés en sciences forensiques » à <https://pbil.univ-lyon1.fr/R/pdf/for7.pdf>.

TABLE 2. Genotype Frequencies in HT and NT Populations

	AA	AB	AC	AD	AE	BB	BC	BD	BE	CC	CD	CE	DD	DE	EE	Total
HT	10	22	13	15	40	14	7	6	22	3	7	18	2	16	32	227
NT	11	12	18	26	23	3	6	18	27	4	9	11	8	14	16	206

All 15 separate genotypes (made up from the 5 possible alleles A-E) were identified from our analyses of TH-STR. There was a significant difference of genotype frequencies between HT and NT cohorts: $P < 0.001$ ($\chi^2 = 36.57$, 14 df).

FIGURE 3 – Copie d'écran de la table 2 de [28]. Les groupes sont notés « HT » pour les 227 individus ayant de l'hypertension et « NT » pour les 206 n'en ayant pas. Les allèles de TH01 sont notés ici de A à E.

```

Pearson's Chi-squared test with simulated p-value (based on 2000
replicates)
data: dta[, 1:15]
X-squared = 35.252, df = NA, p-value = 0.0009995

```

Le test du χ^2 nous permet de retrouver les résultats de la figure 3 page 18 : avec un risque de première espèce de 5 %, les données expérimentales sont en contradiction avec l'hypothèse d'indépendance entre les génotypes et la présence d'hypertension. Calculons les fréquences alléliques dans les deux groupes puis testons si elles diffèrent :

```

dta$A <- with(dta, 2*AA + AB + AC + AD + AE)
dta$B <- with(dta, 2*BB + AB + BC + BD + BE)
dta$C <- with(dta, 2*CC + AC + BC + CD + CE)
dta$D <- with(dta, 2*DD + AD + BD + CD + DE)
dta$E <- with(dta, 2*EE + AE + BE + CE + DE)
t(dta[, 17:21])
  HT NT
A 110 101
B  85  69
C  51  52
D  48  83
E 160 107

round(t(dta[, 17:21]/rowSums(dta[, 17:21])), 2)
  HT NT
A 0.24 0.25
B 0.19 0.17
C 0.11 0.13
D 0.11 0.20
E 0.35 0.26

chisq.test(dta[, 17:21])
Pearson's Chi-squared test
data: dta[, 17:21]
X-squared = 19.938, df = 4, p-value = 0.0005138

```

On retrouve ici les résultats de [28] donnés dans la figure 4 page 19 : avec un risque de première espèce de 5 %, les données expérimentales sont en contradiction avec l'hypothèse d'indépendance entre les allèles et la présence d'hypertension. Une représentation graphique de type « *mosaicplot* » [14, 11, 12] permet de visualiser facilement les données :

```

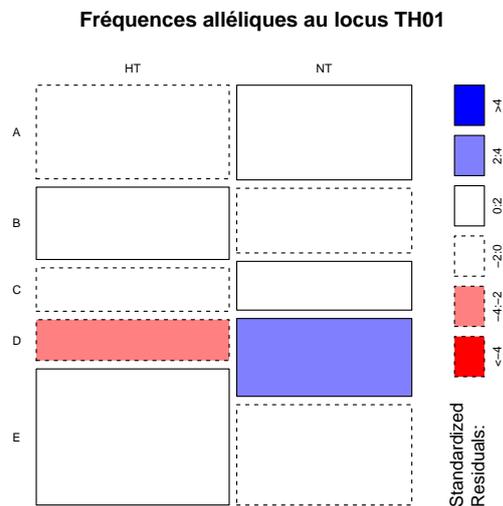
mosaicplot(dta[,17:21], shade =TRUE, las = 1,
main = "Fréquences alléliques au locus TH01")

```

TH Allele	NT	HT
A	0.24	0.24
B	0.17	0.19
C	0.13	0.11
D	0.20	0.11
E	0.26	0.35

Five alleles (A–E) were identified from our analysis of the TH-STR. A highly significant difference in allele frequencies was found between HT and NT cohorts: $P < 0.0005$ ($\chi^2 = 19.9$, 4 *df*). This chromosomal χ^2 analysis is an even more powerful test in an association model compared with genotype χ^2 analysis (see text).

FIGURE 4 – Copie d'écran de la table 3 de [28] donnant les fréquences alléliques du locus TH01 pour les individus ayant de l'hypertension et ceux n'en ayant pas.

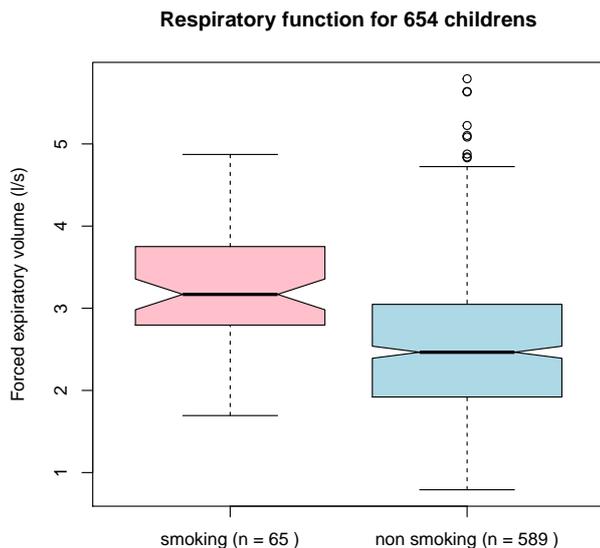


C'EST donc l'allèle D qui explique pourquoi le test d'indépendance du χ^2 est significatif : les individus hypertendus ont tendance à être moins souvent porteurs de cet allèle que ceux qui ne le sont pas. Mais attention : cette association est de nature probabiliste et non déterministe. Ce n'est pas parce que l'on est porteur de l'allèle D que l'on est automatiquement non-hypertendu (il y a 46 individus porteurs de l'allèle D qui sont hypertendus). Ce n'est pas parce que l'on n'est pas porteur de l'allèle D que l'on est automatiquement hypertendu (il y a 131 individus non porteurs de l'allèle D qui ne sont pas hypertendus).

UNE autre difficulté commune à toutes les études de recherche d'association entre un allèle et une maladie est que de nombreux facteurs, dits de nuisance, peuvent perturber gravement l'interprétation des résultats. Voici un exemple célèbre de facteur de nuisance :

```
fev <- read.table("http://pbil.univ-lyon1.fr/members/lobry/corcau/fev.txt")
names(fev) <-c("seqnbr", "subjid", "age", "fev", "height", "sex", "smoke")
smk <- fev[fev$smoke == 1, ] # smokers
```

```
nsmk <- fev[fev$smoke == 2, ] # non smoker
boxplot(fev$fev~fev$smoke, col = c("pink","lightblue"),
        names = c(paste("smoking (n =", nrow(sm), ")"),
                  paste("non smoking (n =", nrow(nsmk), ")")),
        notch = TRUE,
        ylab = "Forced expiratory volume (l/s)",
        main = paste("Respiratory function for", nrow(fev),"childrens"))
```



CETTE représentation graphique permet de comparer la capacité expiratoire chez 654 enfants classés en deux groupes selon qu'ils sont fumeurs ou non. Un test de comparaison de moyennes nous montrerait qu'il y a une différence très significative entre les deux groupes mais le problème ici est que la capacité expiratoire est meilleure chez les enfants fumeurs. Fumer n'est évidemment pas bon pour les enfants. Le facteur de nuisance que nous avons négligé ici est l'âge des enfants : la capacité expiratoire augmente avec la croissance des enfants et comme les fumeurs sont plus âgés on biaise les résultats. Pankaj SHARMA *et al.* [28] ont bien entendu fait attention à constituer des groupes homogènes pour les facteurs de nuisance connus pour moduler l'hypertension (*viz.* sexe, âge, indice de masse corporelle, consommation de cigarettes et d'alcool, taux de cholestérol), mais rien ne peut garantir que l'on n'a pas oublié un facteur de nuisance trivial.

EN Allemagne, d'après [15] : « [a] recent order of the German Federal Constitutional Court¹⁴ stated that it is “decisive ... (to ensure that a DNA investigation is not in conflict with the civil rights of a citizen) ... that conclusions concerning the personality, like ... diseases ... are not permitted”. However, in clinical studies, a positive association between TH01 alleles 9.3 and 10 and essential hypertension (HT) [28] was reported. »

14. Order of the Federal Constitutional Court of Germany (2BvR1741/999, Dec. 14th 2000), *NeueJuristische Wochenschrift* (12) 879 :882.

TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys
TTC	Phe	TCC	Ser	TAC	Tyr	TGC	Cys
TTA	Leu	TCA	Ser	TAA	Stp	TGA	Stp
TTG	Leu	TCG	Ser	TAG	Stp	TGG	Trp
CTT	Leu	CCT	Pro	CAT	His	CGT	Arg
CTC	Leu	CCC	Pro	CAC	His	CGC	Arg
CTA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CTG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser
ATC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
ATA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
ATG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly
GTC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GTA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GTG	Val	GCG	Ala	GAG	Glu	GGG	Gly

TABLE 3 – Le code génétique standard.

2.2.3 Les références aux résolutions du Conseil de l'Union européenne

Lors de la discussion du jeudi 16 janvier 2003 le rapporteur : « Monsieur BLISKO, le casier judiciaire et le FNAEG n'ont rien à voir. Vous n'obtiendrez jamais la moindre information du FNAEG à partir d'un nom, à la différence du casier judiciaire. M. LE ROUX le sait aussi bien que Gérard LÉONARD et que tous ceux qui, à Écully, ont rencontré les agents du FNAEG et constaté leur admirable travail. Rien, dans le FNAEG, n'est "codant". Vous ne pourrez jamais, à partir d'un prélèvement d'ADN, tracer le portrait-robot de qui que ce soit. »

La rédaction est inchangée dans le projet de loi adopté par l'Assemblée nationale, en première lecture, le mardi 28 janvier 2003.

La Commission mixte paritaire a adopté l'article 15 dans le texte de l'Assemblée nationale le 4 février 2003.

La rédaction est inchangée dans le projet de loi adopté par l'Assemblée nationale le 12 février 2003 et par le Sénat le 13 février 2003.



FIGURE 5 – Copie d'écran (2015-10-14) d'une partie de l'article de la version anglaise de Wikipédia consacré aux didones (une famille typographique ancêtre de la police de caractères *Computer modern* de Donald KNUTH utilisée par défaut dans les documents scientifiques composés avec L^AT_EX). Notez l'emploi de la majuscule dans « Contenant le texte du **C**ode civil » sous le « TOME PREMIER » .

Références

- [1] C. Alves, L. Gusmão, and A. Amorim. STR data (AmpF ℓ STR Profiler Plus and GenePrint CTTv) from Mozambique. *Forensic Science International*, 119 :131–133, 2001.
- [2] Anonymous. *AmpF ℓ STR[®] Identifiler[®] PCR Amplification Kit. User's Manual*. Applied Biosystems, Foster City, CA, USA, 2006. PN 4323291D.
- [3] W. Bar, B. Brinkmann, P. Lincoln, W.R. Mayr, and U. Rossi. DNA recommendations. 1994 report concerning further recommendations of the DNA commission of the ISFH regarding PCR-based polymorphisms in STR (short tandem repeat) systems. *Int. J. Leg. Med.*, 107 :159–160, 1994.
- [4] G. Bousquet. *Autobiografic*. Pascal Galodé, 18, rue de Toulouse, 35400 Saint-Malo, France, 2010.
- [5] A. Buquet. *Manuel de criminalistique moderne et de police scientifique (4^e édition)*. Presses Universitaires de France, 6, avenue Reille, 75014 Paris, France, 2008.
- [6] D. Charif and J.R. Lobry. SeqinR 1.0-2 : a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In H.E. Roman U. Bastolla, M. Porto and M. Vendruscolo, editors, *Structural approaches to sequence evolution : Molecules, networks, populations*, Biological and Medical Physics, Biomedical Engineering, pages 207–232. Springer Verlag, New York, USA, 2007. ISBN 978-3-540-35305-8.
- [7] J. Commaille. *À quoi nous sert le droit ?* Gallimard, 5, rue Gaston-Gallimard, 75328 Paris Cedex 07, France, 2015.
- [8] G. Cornu. *Vocabulaire juridique (10^e édition)*. Presses Universitaires de France, 6, avenue Reille, 75014 Paris, France, 2014.
- [9] J. Dallest. *Mes homicides*. Robert Laffont, 30, place d'Italie, 75013 Paris, France, 2015.
- [10] F. Dupuch. *Nos "experts"*. Michel Lafon, 118, avenue Achille Peretti, 92521 Neuilly-sur-Seine Cedex, France, 2010.
- [11] J. W. Emerson. Mosaic displays in S-PLUS : a general implementation and a case study. *Statistical Computing and Graphics Newsletter (ASA)*, 9 :17–23, 1998.
- [12] M. Friendly. Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association*, 89 :190–200, 1994.
- [13] C. Gavat. *96 heures*. Michalon, 9, rue de l'École-Polytechnique, 75005 Paris, France, 2013.
- [14] J.A. Hartigan and B. Kleiner. A mosaic of television ratings. *The American Statistician*, 38 :32–35, 1984.
- [15] M. Klintschar, U. Immel, D. Stiller, and M. Kleiber. HumTH01 and blood pressure. an obstacle for forensic applications? *International Congress Series*, 1261 :589–591, 2004.

- [16] J.V. Maizel and R.P. Lenk. Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proceedings of the national academy of sciences of the United States of America*, 78 :7665–7669, 1981.
- [17] P. Malinvaud. *Introduction à l'étude du droit (15^e édition)*. LexisNexis, 141, Rue de Javel, 75015 Paris, France, 2015.
- [18] A. Mannucci, K.M. Sullivan, P.L. Ivanov, and P. Gill. Forensic application of a rapid and quantitative DNA sex test by amplification of the X-Y homologous gene amelogenin. *International Journal of Legal Medicine*, 106 :190–193, 1994.
- [19] Y. Mayaud and C. Gayet. *Code pénal (108^e édition)*. Dalloz, 31-35, rue Froidevaux, 75685 Paris Cedex 14, France, 2011.
- [20] M. Monteil. *Flic, tout simplement*. Michel Lafon, 118, avenue Achille Peretti, 92521 Neuilly-sur-Seine Cedex, France, 2008.
- [21] Y. Nakahori, O. Takenaka, and Y. Nakagome. A human X-Y homologous region encodes "amelogenin". *Genomics*, 9 :264–269, 1991.
- [22] F.-J. Pansier. *Méthodologie du droit (7^e édition)*. LexisNexis, 141, Rue de Javel, 75015 Paris, France, 2016.
- [23] H. Pelletier. *Code pénal (28^e édition)*. LexisNexis, 141, Rue de Javel, 75015 Paris, France, 2016.
- [24] P. Ramaël. *Hors procédure*. Calmann-Lévy, 21, rue du Montparnasse, 75298 Paris Cedex 06, France, 2015.
- [25] J.-F. Renucci and C. Gayet. *Code de procédure pénale (50^e édition)*. Dalloz, 31-35, rue Froidevaux, 75685 Paris Cedex 14, France, 2009.
- [26] J.-P. Royer, N. Derasse, J.-P. Alline, B. Durand, and J.-P. Jean. *Histoire de la justice en France du XVII^e siècle à nos jours (5^e édition)*. Presses Universitaires de France, 6, avenue Reille, 75014 Paris, France, 2016.
- [27] M. Sapanet. *Chroniques d'un médecin légiste*. Jean-Claude Gawsewitch, 130, rue de Rivoli, 75001 Paris, France, 2009.
- [28] P. Sharma, A. Hingorani, H. Jia, M. Ashby, R. Hopper, D. Clayton, and M.J. Brown. Positive association of tyrosine hydroxylase microsatellite marker to essential hypertension. *Journal of Statistical Software*, 32 :676–682, 1998.
- [29] K. M. Sullivan, C. P. Mannucci, A. and Kimpton, and P. Gill. A rapid and quantitative DNA sex test : fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *BioTechniques*, 15 :637–641, 1993.
- [30] F. Terré. *Introduction générale au droit (10^e édition)*. Dalloz, 31-35, rue Froidevaux, 75685 Paris Cedex 14, France, 2015.
- [31] M. Trévidic. *Au cœur de l'antiterrorisme*. Jean-Claude Lattès, 17, rue Jacob, 75006 Paris, France, 2011.