

10581



Institut français de recherche scientifique
pour le développement en coopération



PHYLOGÉNIES MOLÉCULAIRES
INTRASPÉCIFIQUES COMPARÉES
ENTRE UN PROCARYOTE
ESCHERICHIA COLI
ET UN PROTOZOAIRE PARASITE
TRYPANOSOMA CRUZI

Marie-Thérèse GATTO

Laboratoire de génétique moléculaire des parasites et des vecteurs
UMR CNRS-ORSTOM n°9926

ORSTOM Documentation



010004521

Marie-Thérèse GATTO

PHYLOGÉNIES MOLÉCULAIRES
INTRASPÉCIFIQUES COMPARÉES
ENTRE UN PROCARYOTE *ESCHERICHIA COLI*
ET UN PROTOZOAIRE PARASITE *TRYPANOSOMA CRUZI*



20 JUIN 1996

Montpellier
1996

Ø52
GLOTRY
GAT



F

Les opinions exprimées dans ce document
n'engagent que la responsabilité de leurs auteurs

Ce travail de recherche a été effectué au centre de recherche Orstom de Montpellier, au sein de l'unité mixte de recherche CNRS-ORSTOM n° 9926 «Génétique moléculaire des parasites et des vecteurs» (U.R. 41 des Maladies infectieuses et parasitaires; département Eau et Santé), sous la direction de Michel TIBAYRENC, directeur de l'unité, et avec l'encadrement de Bruno OURY et de Christian BARNABÉ, pour l'obtention par Marie-Thérèse GATTO du Diplôme d'Études Approfondies d'Évolution et Écologie de l'Université des sciences et techniques du Languedoc-Montpellier II, le 11 septembre 1995.

UNIVERSITE MONTPELLIER II SCIENCES ET TECHNIQUES DU LANGUEDOC
INSTITUT NATIONAL AGRONOMIQUE DE PARIS-GRIGNON
ECOLE NATIONALE SUPERIEURE D'AGRONOMIE DE MONTPELLIER
ECOLE PRATIQUE DES HAUTES ETUDES

Diplôme d'études approfondies (DEA)
EVOLUTION ET ECOLOGIE

1995

Marie-Thérèse GATTO

**Phylogénies moléculaires intraspécifiques comparées
entre
un procaryote *Escherichia coli*
et
un protozoaire parasite *Trypanosoma cruzi***

Stage principal effectué dans le laboratoire de *Génétique Moléculaire des Parasites et des Vecteurs*. UMR CNRS-ORSTOM n°9926. Centre de recherche ORSTOM (Montpellier)

Sous la direction du Dr Michel TIBAYRENC

Soutenu le 11 septembre 1995 devant la commission d'examen :

M. L. THALER, Président, Mme I. OLIVIERI, MM. F. BONHOMME, A. CHARRIER,
B. DELAY, J. ESCARRE, J.D. LEBRETON, J. LEPART, J. MAILLET, D. MC KEY,
M. RAMBAL, M. RAYMOND, J. ROY

TABLE DES MATIERES

Résumé	
Abstract	
I Introduction	1
II Généralités	1
III Matériel et méthodes	4
3.1 Matériel biologique	4
3.1.1 <i>Trypanosoma cruzi</i>	4
3.1.2. <i>Escherichia coli</i>	4
3.2 Culture de bactéries	4
3.3 Electrophorèse d'isoenzymes multilocus	4
3.4 Les RAPD	6
3.4.1 Purification de l'ADN	6
3.4.2 La technique RAPD	7
3.5 Méthodes d'analyse des résultats	8
3.5.1 Distances et dendrogrammes	8
3.5.2 Les tests statistiques	9
IV Résultats	11
4.1 <i>Escherichia coli</i>	11
4.1.1 Les isoenzymes	11
4.1.2 Les RAPD	11
4.1.3 Distances génétiques, diversité génétique moyenne	11
4.1.4 Analyse phénétique : les dendrogrammes	12
4.1.5 Les tests statistiques	13
4.2 <i>Trypanosoma cruzi</i>	13
4.2.1 Distances et diversités génétiques	13
4.2.2 Les dendrogrammes	14
4.2.3 Les tests statistiques	14
V Discussion	15
VI Conclusion	20
Références bibliographiques	

Résumé

Le présent travail a comparé la diversité génétique d'un procaryote, *Escherichia coli* et d'un microorganisme eucaryote *Trypanosoma cruzi*, protozoaire parasite agent de la maladie de Chagas.

L'étude a confirmé que les populations d'*Escherichia coli* présentent un fort déséquilibre de liaison, non imputable à la distance géographique. Ceci constitue un argument indirect classique pour une structure clonale, et confirme les hypothèses précédemment proposées par d'autres auteurs.

La comparaison directe a montré que les populations d'*E. coli* apparaissent cependant moins structurées que celles de *T. cruzi*, comme le suggère la disparition du déséquilibre de liaison quand on considère chaque groupe isolément, ou quand on élimine les génotypes répétés de l'analyse.

L'analyse phylogénétique d'*E. coli* recoupe dans ses grandes lignes celle présentée par d'autres auteurs. Il est à noter que la distance de migration séparant les électromorphes est constamment moindre chez *E. coli* que chez *T. cruzi*, ce qui suggère des niveaux de divergence phylogénétique inférieurs.

D'une manière générale, l'approche comparative directe conduit à minimiser les similitudes évolutives entre *E. coli* et *T. cruzi* suggérées par l'analyse des seules données de la littérature.

Le présent travail doit être considéré comme préliminaire, et devra être complété par des analyses plus poussées reposant sur des techniques davantage perfectionnées.

Abstract

The present study compared the genetic diversity of a prokaryotic microorganism, *Escherichia coli*, and a eukaryotic microorganism, *Trypanosoma cruzi*, a parasitic protozoan responsible for Chagas' disease.

The long-term goal of this study, in agreement with the new programme of the guest laboratory, is to try to bring out the common laws that drive the molecular evolution of microorganisms, while highlighting the specificities of each species under study. Towards reaching this goal, a direct comparative approach is required, that relies on the same techniques (here : cellulose acetate electrophoresis and RAPD) and the same statistics.

Our study confirmed that *Escherichia coli* populations exhibit a strong linkage disequilibrium that cannot be explained by geographical distance. This is a classical circumstantial evidence for a clonal structure, and confirms the hypotheses previously proposed by others.

The direct comparison showed that *E. coli* populations appear nevertheless as less structured than *T. cruzi* populations, as suggested by the disappearing of linkage disequilibrium when each cluster is analysed separately, or when repeated genotypes are discarded from the analysis.

Phylogenetic analysis of *E. coli* confirms in the main the one presented by other authors. It is worth noting that migration distance that separates the electromorphs is constantly lower in *E. coli* than in *T. cruzi*, which suggests lower levels of phylogenetic divergence.

Generally speaking, the direct comparative approach leads to minimize the evolutionary similarities between *E. coli* et *T. cruzi* that were suggested by the analysis of published data alone.

The present work should be considered as preliminary, and will have to be completed by more advanced studies that rely upon more perfected techniques.

I INTRODUCTION

La génétique évolutive a trois retombées principales en recherche appliquée (Tibayrenc 1995) :

- le typage permettant le suivi épidémiologique (l'évaluation de la stabilité des génotypes dans l'espace et dans le temps, et l'importance des échanges génétiques)
- la taxonomie
- des études d'aval (comme l'évaluation de l'impact de la diversité génétique des microorganismes sur leurs propriétés biologiques d'importance médicale, telles que la virulence et la résistance aux drogues) (voir table 1 Tibayrenc 1995 Adv. Parasitol. en Annexe O).

Les applications potentielles de la génétique évolutive étant exactement les mêmes quel que soit le microorganisme considéré, nous avons effectué une étude visant à établir des bases pour la comparaison directe de la structure des populations de microorganismes différents. Ainsi par exemple, si l'on s'appuie sur un critère de divergence phylogénétique pour mieux définir les espèces microbiennes, les mêmes échelles de divergences phylogénétiques seront utilisées pour les autres microorganismes.

Le but de cette approche comparative est de dégager les éventuelles lois communes régissant l'évolution moléculaire de ces microorganismes, ainsi que les particularités propres à chacun d'entre eux (Tibayrenc et Ayala 1991, Tibayrenc 1995).

Notre choix s'est porté sur *Trypanosoma cruzi*, protozoaire parasite, agent de la maladie de Chagas et sur la bactérie *Escherichia coli*, deux microorganismes qui ont été très étudiés en génétique des populations. Jusqu'à présent aucune étude comparative directe n'a été réalisée entre ces deux espèces si éloignées sur le plan évolutif. De plus, l'étude de leur polymorphisme génétique a été abordée par des techniques moléculaires sensiblement différentes. C'est pourquoi, pour réaliser notre approche comparative de la structure des populations de ces microorganismes, nous nous sommes efforcés d'utiliser des techniques similaires telles que les RAPD « Random Amplified Polymorphic DNA » et les électrophorèses isoenzymatiques multilocus sur acétate de cellulose (en prenant un maximum d'amorces et de locus enzymatiques en commun). Les données obtenues avec ces deux séries de marqueurs ont été traitées avec les mêmes tests statistiques et les mêmes concepts de génétique des populations.

II GENERALITES

Trypanosoma cruzi est un protozoaire parasite de la famille des Kinétoplastidés, agent de la maladie de Chagas. Cette endémie transmise par des insectes hématophages de la famille des Réduviidés et de la sous-famille de Triatominés est un problème de santé majeur en Amérique Latine car elle affecte 16 à 18 millions de personnes et causerait la mort de 45 000

personnes par an. Chez les enfants, elle est à l'origine de myocardites et de méningoencéphalites fatales à 50 %, seulement deux à trois semaines après l'infection. Cependant les plus grandes complications causées par cette maladie se manifestent dix à vingt ans plus tard chez le tiers des personnes infectées. Cela inclut des cardiopathies chroniques (chez plus de 27 % des personnes infectées), et des désordres digestifs (pour 6 %) et neurologiques chroniques (pour 3 %).

Comme pour les autres microorganismes pathogènes, les études évolutives chez *T. cruzi* présentent de nombreuses retombées potentielles (Annexe 0). Les principaux résultats obtenus dans le cas de l'agent de la maladie de Chagas sont les suivants :

- Tibayrenc *et al.* ont montré que *T. cruzi* avait une structure clonale caractérisée par un fort déséquilibre de liaison (association non aléatoire des génotypes observés à différents locus) et l'existence de génotypes répétés, retrouvés sur de vastes aires géographiques, pendant de longues périodes de temps.
- La structure clonale n'implique pas une absence totale de la recombinaison génétique. Elle indique seulement que la recombinaison est trop rare pour empêcher la propagation de clones stables dans l'espace et dans le temps. Cette stabilité des génotypes rend possible leur utilisation comme marqueurs épidémiologiques. Ce ne serait pas le cas si les génotypes étaient modifiés à chaque génération par la recombinaison.

Des études de génétique des populations concernant la bactérie *E. coli* ont montré une structure clonale comparable à celle de *T. cruzi* avec un fort déséquilibre de liaison et l'existence de génotypes stables dans l'espace et dans le temps (Selander *et al.* 1986).

Escherichia coli peut être un pathogène opportuniste pour l'homme et les animaux. Chez l'homme, cette bactérie provoque dans certains cas des diarrhées, des infections de l'appareil digestif et urinaire, des méningites néonatales et des infections nosocomiales chirurgicales de la peau et de l'appareil urinaire (Hartl et Dykhuizen 1984).

De nombreuses techniques ont été employées pour détecter les variations génétiques au sein des isolats d'*E. coli*, et établir des critères de classification et d'identification pour différencier les souches pathogènes des souches non-pathogènes ; il s'agit par exemple de l'utilisation des sérotypes révélant plus de 1000 combinaisons d'antigènes des groupes O, K, H au sein des isolats d'*E. coli* (Eveland *et al.* 1971), des RFLP « restriction fragment length polymorphism » (Anilionis et Riley 1980, Harshman et Riley 1980), et du séquençage de l'ADN (Milkman et Crawford 1983). Ces techniques ont révélé une variabilité génétique très importante chez *E. coli* mais n'ont pas permis de répondre à un certain nombre de questions de génétique des populations. Les électrophorèses d'isoenzymes multilocus (Ochman *et al.* 1983), développées depuis les années 60 et plus récemment, les RAPD se sont révélés être de très bons outils moléculaires pour étudier la variabilité génétique et la structure des populations de cette espèce (Desjardins *et al.* 1995, Cavé *et al.* 1994). Les isoenzymes sur acétate de cellulose n'avaient, jusqu'à présent, jamais été employées sur *E. coli*.

L'objectif de notre étude n'est donc pas d'établir la position phylogénétique de *T. cruzi* et d'*E. coli* l'un par rapport à l'autre, mais de comparer les tendances évolutives intraspécifiques de ces deux microorganismes avec les mêmes outils techniques et conceptuels.

III MATERIEL ET METHODES

3.1 Matériel biologique

3.1.1 *Trypanosoma cruzi*

L'étude de *T. cruzi* s'est appuyée sur les données de Tibayrenc *et al.*(1993) qui portaient sur la caractérisation génétique de 24 stocks sur la base de 22 loci enzymatiques et de 6 amorces RAPD (Kit A d'Operon Technologies). Les stocks étudiés sont isolés de divers hôtes d'origine géographique différente. Ils sont représentatifs de la variabilité globale de l'espèce *T. cruzi*.

3.1.2 *Escherichia coli*

29 stocks d'*E. coli* ont été sélectionnés parmi 72 stocks de référence (Ochman *et al.* 1984), représentatifs de la diversité génétique de l'espèce. Ils sont d'origine géographique différente et isolés chez des hôtes variés (homme, animaux). Ochman *et al.* (1984) ont montré par une analyse en composantes principales que les 72 stocks de référence se distribuent en trois principaux groupes. Les 29 stocks caractérisés dans notre étude ont été choisis sur la base de ces trois groupes, mais aussi sur leur origine géographique et l'hôte chez lequel ils ont été isolés.(Annexe 10)

3.2 Culture des bactéries

A partir des cryostabilats, les stocks de référence ont été clonés deux fois successivement par étalement sur milieu de culture SOB solide (Tryptone 20/l, extrait de levure 5g/l, NaCl 10 mM, Kcl 2.5 Mm, MgCl₂ 10 mM, MgSO₄ 10 mM). Une préculture de 60 ml de milieu SOB liquide estensemencée avec une colonie et incubée à 37°C sous agitation pendant 4 heures pour atteindre la fin de la phase exponentielle de croissance (DO 550 nm > 0,6). 45 ml sont prélevés pour ensemencer la culture finale de 300 ml (37°C, 4 h). Après centrifugation à 2400 x g, pendant 30 min à 4°C, les culots sont lavés trois fois dans le tampon PBS (Phosphate buffered saline: Na₂HPO₄ 8.1 mM, NaH₂PO₄ 1.9mM, NaCl 0.15 M, pH 7.4) par des centrifugations successives de 10 min à 3200 x g à 4°C. Les culots bactériens sont répartis en tubes de 1,5 ml à sec et conservés à -70°C jusqu'à leur utilisation.

3.3 Electrophorèse d'isoenzymes

Les enzymes, soumises à un champ électrique, migrent à une vitesse qui dépend de leur charge électrique globale. Cette dernière est le résultante des charges individuelles portées par les acides aminés qui constituent ces enzymes. La migration électrophorétique est donc le reflet de la structure primaire qui est directement codée par les gènes. L'électrophorèse enzymatique constitue ainsi une méthode simple, rapide et peu coûteuse pour l'étude de la variabilité génétique.

-Préparation des extraits isoenzymatiques

Les culots sont décongelés et resuspendus dans un volume de solution stabilisatrice d'enzymes (100 ml d'eau distillée, 10 mg NADP, 0,1% B-mercaptoethanol, 0,1% de détergent TRITON X 100) correspondant au poids du culot (1 µl de solution par mg de culot). Ils sont ensuite agités pendant 2,5 min et incubés dans la glace pendant 20 min afin de lyser les cellules. Après centrifugation à 12000 xg pendant 10 min à 4°C, le surnageant, contenant les enzymes hydrosolubles, est récupéré et conservé à -70°C (Ben Abderrazak *et al.* 1993).

- Les systèmes enzymatiques étudiés

10 systèmes enzymatiques ont été sélectionnés parmi 30 testés: GAPD, GOT, G6PD, GPI, MDH, MPI, NHI, NSP, 6PGD, SOD (annexe 1).Le système NHI permettant de montrer l'activité de deux locus génétiques différents, 11 loci au total ont pu être étudiés.

- Support, Migration, Révélation

Les extraits enzymatiques sont déposés à l'aide d'un peigne spécial sur des plaques d'acétate de cellulose préalablement imbibées dans le tampon d'électrophorèse (plus ou moins dilué selon les systèmes: voir annexe 1) pendant 20 min (Laboratoire HELENA). Dans le cas de l'acétate de cellulose, la vitesse de migration dépend principalement de la charge protéique globale. La taille et la forme de la molécule ont peu d'influence sur la vitesse de migration.

Les plaques sont ensuite disposées dans les cuves d'électrophorèse contenant le tampon de migration spécifique à chaque système enzymatique, la zone de dépôt orientée du côté de la cathode, exceptée pour la SOD, pour laquelle le dépôt est central. Le voltage appliqué et le temps de migration sont spécifiques du système enzymatique considéré (voir annexe 1). Des batonnets de glace, refroidis préalablement à -20°C, sont déposés dans les compartiments centraux de la cuve pour éviter un échauffement du système. Des bandelettes de papier filtre assurent la liaison entre la plaque d'acétate de cellulose et le tampon de migration.

Les solutions de révélation, spécifiques de chaque système enzymatique, sont préparées suivant les recettes données en annexe 2. Les activités enzymatiques sont révélées à 37°C sur l'agarose 1,2% contenant les facteurs nécessaires à leur révélation. Les réactions enzymatiques sont bloquées en plongeant les plaques dans l'acide acétique 5%. Après un rinçage à l'eau, celles-ci sont séchées à température ambiante.

-Lecture des profils isoenzymatiques

Les profils isoenzymatiques (ou zymogrammes) sont lus locus par locus. Une interprétation allélique n'étant pas toujours réalisable pour tous les systèmes enzymatiques étudiés, nous avons effectué une lecture génotypique. Chaque stock sera ainsi caractérisé, pour chaque locus, par un certain nombre de bandes. La bande la plus rapide sera notée 1, la deuxième, plus lente, sera notée 2, et ainsi de suite.

Les génotypes lus à chaque locus constituent des génotypes élémentaires, dont la combinaison constitue le génotype multilocus du stock à l'étude. Dans l'hypothèse nulle de panmixie, la fréquence attendue d'un génotype multilocus est le produit de la fréquence observée des génotypes élémentaires qui le composent (Tibayrenc *et al.* 1990, Tibayrenc 1995).

Tous les stocks présentant le même profil isoenzymatique multilocus, indépendamment de toute interprétation génétique, sont rangés dans le même zymodème. Ce dernier terme est celui qui est le plus souvent employé dans la littérature parasitologique. Dans la littérature bactériologique, lui correspond le terme de "type électrophorétique" (electrophoretic type = ET).

3.4 Les RAPD

Développés depuis quelques années maintenant (Welsh et McClelland 1990, Williams *et al.* 1990), les RAPD se révèlent être de très bons marqueurs génétiques, utilisables en génétique des populations. Le principe de cette méthode est l'amplification d'ADN avec des amorces (en général décamerique), dont la séquence est choisie de façon arbitraire. Elles s'apparient aux séquences complémentaires sur les deux brins d'ADN dénaturé. La Taq polymérase polymérise les brins complémentaires, et le fragment d'ADN situé entre les 2 sites de fixation de l'amorce, est synthétisé. La réaction génère ainsi un profil d'amplification de fragments de tailles variables pouvant être analysés sur gel d'agarose.

Les RAPD permettent de détecter un polymorphisme engendré par divers événements mutationnels comme des délétions et des insertions, ce qui modifie la taille du segment amplifié ou le site d'appariement avec l'amorce. Ces remaniements génétiques, étant la plupart du temps irréversibles, les RAPD constituent d'excellents marqueurs génétiques pour l'identification génétique des stocks.

3.4.1 Purification de l'ADN

-Extraction de l'ADN

L'extraction de l'ADN génomique comporte trois étapes:

-Lyse des parois, des cellules et hydrolyse de l'ARN:

Le culot (100 mg) est suspendu dans 1 ml de Tampon de lyse (Tris-HCl pH 8 50 mM, EDTA 100 mM, Lysosyme 1 mg/ml). Après incubation à 37°C pendant 15 min en présence de lysosyme (1mg/ml) pour détruire les parois bactériennes, les cellules sont lysées sous l'action détergente du SDS (0,5 %). L'ARN bactérien est ensuite éliminé par incubation en présence de Rnase A (100 µl/ml) à 37°C pendant 1h30.

-Déprotéinisation:

Les protéines sont digérées après incubation en présence de Protéinase K (200 µl/ml) pendant une nuit à 50°C et éliminées par au moins 3 extractions au phénol suivies de deux extractions au chloroforme et de trois extractions à l'éther. Le phénol dénature et précipite les protéines. Le chloroforme extrait les lipides et élimine les traces de phénol. L'éther élimine les traces de chloroforme. Après chaque extraction, la solution est bien homogénéisée et

centrifugée à 10000 xg pendant 5 à 10 min à 4°C. Seule la phase aqueuse contenant l'ADN est récupérée.

-Précipitation de l'ADN:

L'ADN extrait est précipité en présence de NaCl 0,1 M et 2,5 volumes d'éthanol absolu froid pendant une nuit à -70°C. Après une centrifugation à 15000 x g pendant 30 min à 4°C, le culot d'ADN est lavé trois fois à l'éthanol 70% froid pour éliminer les sels, et resuspendu dans un volume minimum d'eau stérile.

-Dosage spectrophotométrique

L'absorption de la solution d'ADN est mesurée entre 210 et 310 nm (UV). La présence de solvant organique et de protéines peut occasionner une distorsion du spectre. Le pic d'absorption maximum d'un ADN purifié se situe entre 257 et 260 nm. La concentration d'ADN est calculée à partir de la densité optique (DO) au pic d'absorption; une unité de DO représente une concentration de 50 µg/ml d'ADN double brin. La solution d'ADN est répartie en aliquots et conservée dans l'EDTA (1 mM) à -20°C.

3.4.2 La technique RAPD

L'ADN est amplifié selon les conditions décrites par Williams *et al.* (1990).

- Réactifs et matériel

Les amorces sont des décarnères dont la séquence est choisie de façon arbitraire et ne présente aucun palindrome. Dans notre étude, parmi les 20 amorces différentes du Kit A d'Operons Technologies testées, douze ont été retenues : A1, A2, A3, A4, A7, A9, A10, A13, A16, A17, A19, A20 (Annexe 2).

Les amplifications ont été réalisées sur un thermocycleur PTC 100 (MJ Research). Le programme d'amplification dure environ six heures et comprend 45 cycles comportant trois étapes : une étape de dénaturation (1 min ; 94°C), une étape d'appariement (1 min ; 36°C), une étape de synthèse (52 min ; 72°C).

-Amplification et visualisation des bandes

Les manipulations sont effectuées dans les conditions les plus stériles possibles pour éviter toute amplification d'ADN contaminant qui pourrait interférer avec nos résultats.

La réaction est effectuée dans 60 µl de milieu réactionnel (Tampon Boehringer mannheim spécifique de la Taq 1x, dNtp 100 mM de chaque, amorce 200 mM, ADN 20 ng, Taq DNA polymérase 15 mU/µl).

-le gel d'électrophorèse

Les fragments d'ADN amplifiés sont séparés selon leur taille par électrophorèse sur gel d'agarose 1,6% en tampon TAE 1X(Tris acétate 0,04 M, EDTA 0,001 M). 20 µl de chaque échantillon amplifié sont mélangés à 2 µl de tampon de charge (solution d'ADN du phage λ 450 ng/µl, enzyme de restriction Bst EII 8 U/µl, Tampon Boehringer 10 x) qui augmente la densité

du dépôt et permet de suivre la migration grâce au bleu de Bromophénol. Un champ électrique est établi à 85 volts de la cathode vers l'anode. A la fin de la migration, le gel est coloré dans un bain de tampon TAE contenant 0,5 µg/ml de Bromure d'éthidium pendant 45 min. Les bandes sont révélées par fluorescence sous lumière UV (254 nm).

-Lecture des profils RAPD

Chaque bande est codée par un nombre (le nombre 1 correspondra à la bande la plus lente ou l'ADN le plus lourd). Chaque stock est ainsi représenté par un ensemble de nombres pour chaque amorce, ce qui constitue un profil RAPD et tous les stocks apparaissant identiques pour les six amorces considérées seront rangés dans un même rapdème (par analogie au zymodème pour les isoenzymes).

Chaque amorce amplifie un nombre inconnu de locus, et il n'est pas certain que pour une amorce donnée, les locus amplifiés soient les mêmes d'un génotype à l'autre. Mais les chances de voir deux amorces différentes révéler le même polymorphisme sont négligeables (McClelland, communication personnelle). Ceci est important, car dans les calculs de déséquilibre de liaison, de réunir ensemble des locus différents ne constitue pas un biais, mais seulement une perte d'information, alors que de distinguer davantage de locus différents qu'il n'y en a effectivement constitue un biais considérable (Tibayrenc 1995).

Pour une amorce donnée, chaque profil RAPD est assimilé à un génotype élémentaire spécifique. Si l'on considère l'ensemble des amorces, comme pour les isoenzymes, la fréquence attendue d'un génotype multilocus défini par la combinaison des génotypes élémentaires obtenus avec les différentes amorces, est le produit de la fréquence observée de chacun des génotypes élémentaires.

3.5 Méthodes d'analyse des résultats

3.5.1 Distances et dendrogrammes

Etant donné que les profils RAPD ne se prêtent pas à une lecture allélique, il nous est impossible d'utiliser la distance génétique standard de Nei pour évaluer les degrés de dissemblances génétiques entre les stocks. La distance de Jaccard ne requiert pas de lecture allélique et permet de quantifier la dissemblance phénétique entre zymodèmes et entre rapdèmes. Les relations entre stocks sont estimées par le pourcentage de bandes communes entre stocks pris 2 à 2. Cette distance se base donc sur le caractère présence-absence des bandes. Elle se calcule selon la formule :

$$D = 1 - (a / (a + b + c))$$

dans laquelle :

a = nombre de bandes communes aux deux génotypes comparés

b = nombre de bandes présentes dans le 1er génotype et absentes dans le second

c = nombre de bandes absentes dans le 1er génotype et présentes dans le second

-Des dendrogrammes UPGMA (Unweight Pair Group Method with Arithmetic Average) ont été élaborés, à partir des matrices de distances de Jaccard, par une classification hiérarchique ascendante qui inclut chaque stock dans un groupe puis lie et hiérarchise ces groupes. Ils constituent un bon résumé visuel des matrices de distance (Sneath et Sokal 1973).

3.5.2 Les tests statistiques

Les deux conséquences fondamentales de la reproduction sexuée sont la ségrégation des allèles à un locus donné et la recombinaison des génotypes d'un locus à l'autre.

Deux catégories de tests existent pour établir la structure des populations : les tests de ségrégation et les tests de recombinaison. Les premiers nécessitent que le niveau de ploïdie de l'organisme considéré soit connu, que l'organisme ne soit pas haploïde et que l'on puisse identifier des allèles ou des locus, ce qui n'est pas le cas pour les seconds (). Or la diploïdie de *T.cruzi* reste une hypothèse de travail (Tibayrenc *et al.* 1981, Lanar *et al.* 1981). *E.coli* est haploïde, et enfin, il est difficile ou impossible d'individualiser des allèles avec les isoenzymes ou les RAPD dans le cas des microorganismes. Pour cela, seuls les tests de recombinaison (recherchant des indices de déséquilibre de liaison) seront utilisés dans le présent travail. Tous ces tests prennent pour hypothèse nulle une situation panmictique dans laquelle les échanges génétiques au sein de la population étudiée se font au hasard entre les individus. Des écarts importants par rapport à la panmixie peuvent donc être considérés comme des arguments indirects (*circumstantial evidence*) en faveur d'une inhibition des échanges géniques et partant, d'une structure clonale. Les biais potentiels apportés par : (i) la séparation géographique ou temporelle (dérive génétique entraînant des différences de fréquences alléliques ou géniques entre populations) ; (ii) la sélection naturelle ; ont été analysés par Tibayrenc *et al.* (1991).

Les tests statistiques utilisés dans cette étude sont les suivants :

- le test d1: évalue la probabilité d'échantillonner un génotype multilocus avec un effectif égal ou supérieur à celui qui est effectivement observé. Ce test est basé sur le fait, qu'en cas de reproduction clonale, les génotypes se reproduisent de façon identique d'une génération à l'autre; certains donc seront représentés en excès par rapport aux prévisions panmictiques.

- le test d2: évalue la probabilité d'observer n'importe quel génotype multilocus avec un effectif supérieur ou égal à celui du génotype le plus abondant dans l'échantillon.

- le test e: évalue la probabilité d'observer un nombre de génotypes différents inférieur ou égal à celui qui est effectivement observé dans l'échantillon. En cas de reproduction asexuée, on s'attend à ce que le nombre total de génotypes soit diminué par rapport à une population sexuée.

- le test f: évalue la probabilité d'observer un niveau de déséquilibre de liaison supérieur ou égal à celui observé dans l'échantillon.

Les tests d2, e, et f sont basés sur des simulations de Monte-Carlo avec 10 000 itérations.

- le test g: repose sur la corrélation observée entre des marqueurs génétiques différents dans une population donnée. Cette corrélation est calculée à partir des matrices de distances RAPD et

isoenzymatiques. Elle est évaluée par un test de corrélation classique ou plus rigoureusement par un test de Mantel (Mantel, 1967). Brièvement, ce test fait appel à une simulation de Monte-Carlo avec 10^4 tirages, qui permute au hasard les différentes cellules d'une des deux matrices. Il ne fait aucune inférence quant au degré de liberté, à la différence d'un test de corrélation classique. Les tests d1, d2, e, f et g sont tirés de Tibayrenc *et al.* (1990).

IV RESULTATS

5.1 *Escherichia coli*

5.1.1 Les isoenzymes

A l'origine, notre étude devait porter sur 10 systèmes enzymatiques et 29 stocks mais les stocks R6 et R30, dont les extraits altérés donnaient des profils non exploitables, ont dû être écartés de notre étude.

Deux systèmes enzymatiques, la GOT et la NSP, apparaissent monomorphes. Pour les locus polymorphes, le nombre d'allèles varie de 2 pour la GAPD à 6 pour la MPI (Fig.2). Le profil de la SOD révèle deux bandes (Annexe 3).

La lecture des profils enzymatiques ne permet pas de distinguer les trois groupes décrits par Ochman *et al.* (1984), seule la GPI met en évidence deux groupes (Fig.1): l'un correspondant au groupe I d'Ochman *et al.* (1984) avec la même composition de stocks et caractérisé par une bande à migration rapide, et l'autre comprenant les 18 souches restantes caractérisé par des bandes à migrations plus lentes.

De très faibles différences de migration existant entre deux électromorphes successifs (de l'ordre du mm) caractérisent les profils enzymatiques de *E. coli* (contrairement à *T. cruzi*).

5.1.2 Les RAPD

Les 20 amorces du Kit A d'Operon Technologies ont été testées sur dix stocks. Aucune amplification n'a été obtenue avec les amorces A5, A8 et A15. Douze amorces ont été sélectionnées pour leur facilité de lecture et testées sur l'ensemble des stocks. Tous ont montré des profils d'amplification (Annexe 4).

Le nombre de bandes générées varie de 2, avec l'amorce A1, à 20 avec l'amorce A4. Toutes les amorces ont généré des profils polymorphes, exceptées les amorces A1 et A2 (Fig.3) qui ont donné des profils simples comportant 2 et 6 bandes respectivement. Trois amorces (A1, A2 et A16) ont permis de mettre en évidence une bande partagée par l'ensemble des stocks : une bande de 650 pb pour l'amorce A1, une de 670 pb pour l'amorce A16 et une de 900 pb pour l'amorce A2. Le profil d'amplification généré par l'amorce A10 (Fig.4) a révélé trois bandes de tailles différentes caractérisant chacune un des trois groupes définis par Ochman *et al.* (1984). Les stocks R52, R54, R57, R58 et R59, appartenant au groupe III d'Ochman *et al.* (1984), partagent sur les profils des amorces A2, A4, A19 et A20 une même bande qui les caractérise.

5.1.3 Distances génétiques, diversité génétique moyenne

La diversité génétique moyenne H correspond à la moyenne arithmétique de la diversité génétique h calculée pour chaque locus ou pour chaque profil d'amplification ($h = 1 - \sum q_i^2$ où q_i est la fréquence du i ème génotype au locus considéré). Les valeurs de diversité génétique

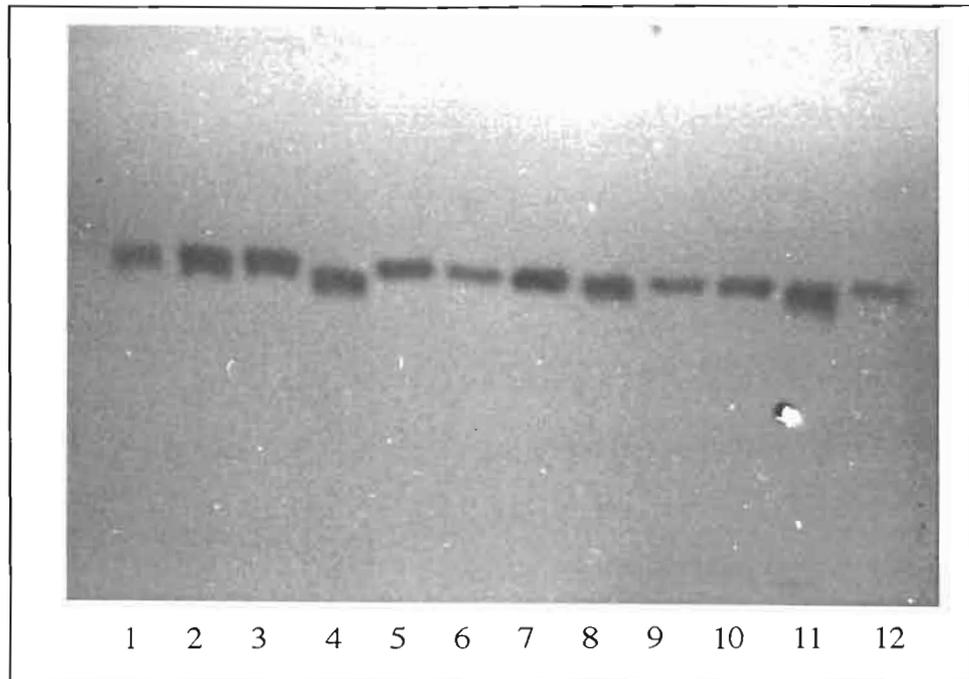


Fig 1 : Profils électrophorétiques obtenus avec la GPI
(1 = R4, 2 = R6, 3 = R9, 4 = R27, 5 = R13, 6 = R14, 7 = R15, 8 = R16,
9 = R52, 10 = R18, 11 = R22, 12 = R25)

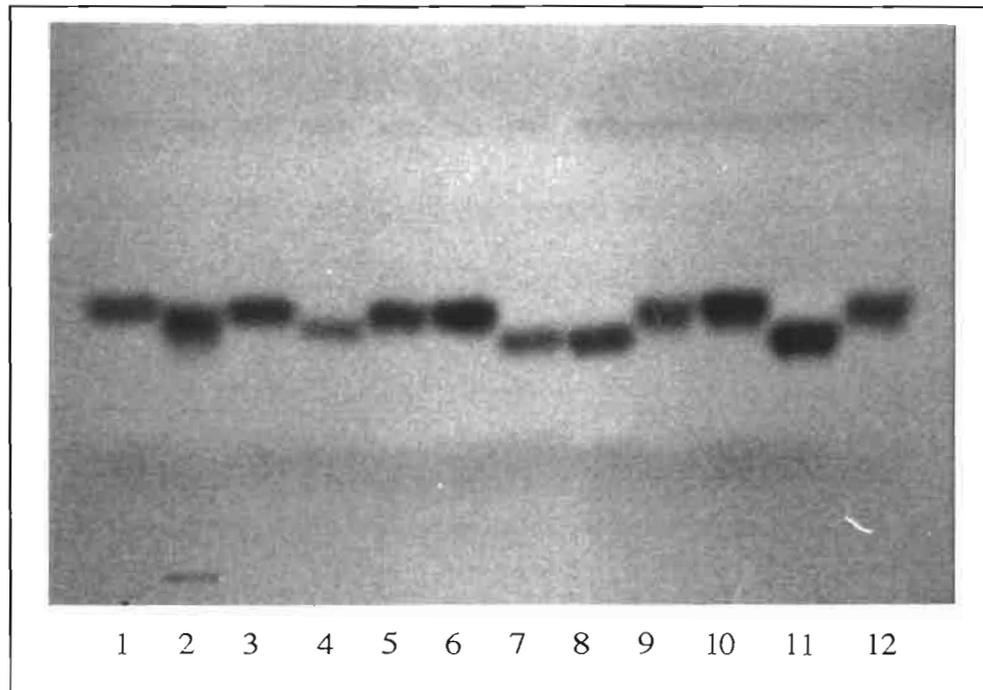


Fig 2 : Profils électrophorétiques obtenus avec la MPI
(1 = R4, 2 = R6, 3 = R9, 4 = R27, 5 = R13, 6 = R14, 7 = R15, 8 = R16,
9 = R52, 10 = R18, 11 = R22, 12 = R25)

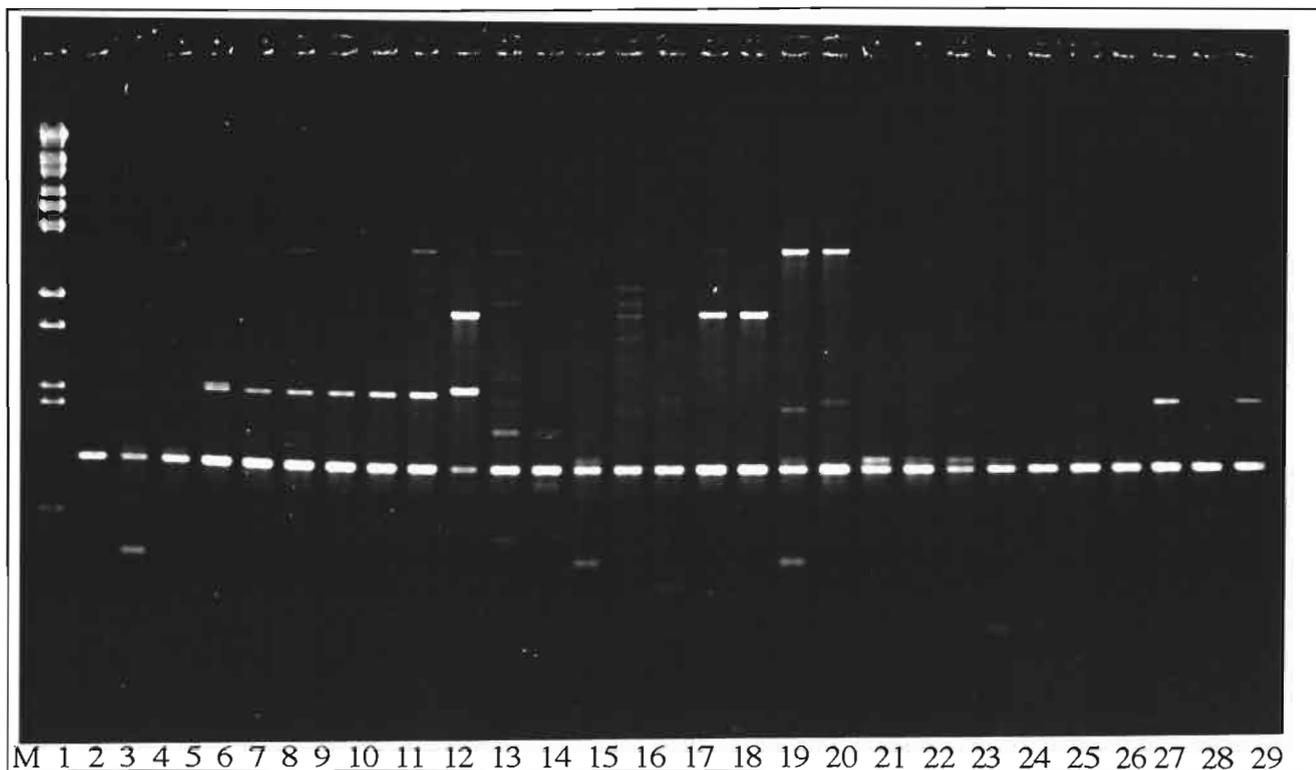


Fig. 3 : Profils d'amplification obtenus avec l'amorce A2
 (M = Marqueur moléculaire, 1 = R4, 2 = R6, 3 = R9, 4 = R13, 5 = R14, 6 = R15, 7 = R16, 8 = R18, 9 = R22, 10 = R25, 11 = R27, 12 = R30, 13 = R36, 14 = R40, 15 = R42, 16 = R43, 17 = R45, 18 = R47, 19 = R48, 20 = R50, 21 = R52, 22 = R54, 23 = R57, 24 = R58, 25 = R59, 26 = R61, 27 = R67, 28 = R68, 29 = R71)

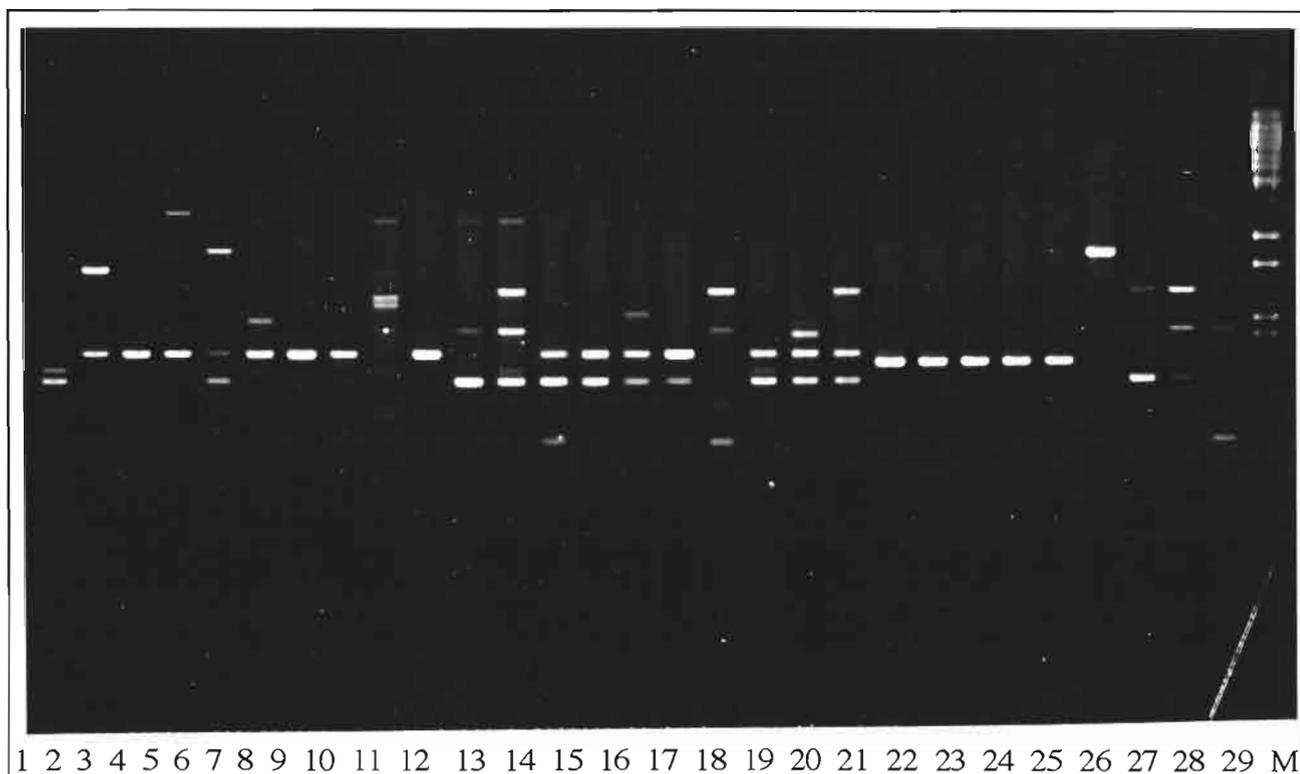


Fig. 4 : Profils d'amplification obtenus avec l'amorce A 10
 (1 = R4, 2 = R6, 3 = R9, 4 = R13, 5 = R14, 6 = R15, 7 = R16, 8 = R18, 9 = R22, 10 = R25, 11 = R27, 12 = R30, 13 = R36, 14 = R40, 15 = R42, 16 = R43, 17 = R45, 18 = R47, 19 = R48, 20 = R50, 21 = R52, 22 = R54, 23 = R57, 24 = R58, 25 = R59, 26 = R61, 27 = R67, 28 = R68, 29 = R71, M = Marqueur moléculaire)

moyenne (H = 0.81) et de distance génétique moyenne (D = 0.65 avec les 12 amorces) obtenues avec les RAPD sont plus grandes que celles obtenues avec les isoenzymes (D = 0.47, H = 0.33). Le nombre moyen de génotypes obtenus avec les données RAPD est trois fois plus élevé que celui obtenu avec les données isoenzymatiques. Les RAPD révèlent un polymorphisme plus important que les isoenzymes (Fig.5).

Les diversités génotypiques (égales au rapport entre le nombre de génotypes différents obtenus avec un marqueur donné, et le nombre de stocks étudiés) sont comparables. Ceci est dû au fait que nous retrouvons presque le même nombre de génotypes répétés avec les deux marqueurs : soit 4 génotypes répétés avec les isoenzymes, et 3 avec les RAPD.

	Distance moyenne	Distance min	Distance max	Distance écart-type	Diversité génétique moy	Diversité génotypique	nb moy de génotypes
<i>Escherichia coli</i>							
nos isoenzymes	0,48	0,09	0,88	0,15	0,34	0,85	3,5
isoenzymes d'Ochman	0,63	0,16	0,9	0,15	0,45	0,88	3,9
RAPD (12 amorces)	0,65	0,09	0,82	0,12	0,81	1	12,58
RAPD combinaison A	0,64	0,14	0,84	0,12	0,75	0,88	11,5
RAPD combinaison B	0,67	0,09	0,91	0,14	0,76	1	12,16
<i>Trypanosoma cruzi</i>							
isoenzymes	0,63	0,07	0,87	0,23	0,59	0,7	5
RAPD	0,75	0,08	1	0,27	0,8	0,75	8,83

Fig. 5 : Indices génétiques calculés avec les deux séries de données RAPD et isoenzymatiques de *E. coli* et *T. cruzi*

5.1.4 Analyse phénétique : les dendrogrammes

Trois dendrogrammes ont été construits à partir des données RAPD : l'un en prenant l'ensemble des douze amorces testées, le second une combinaison A de 6 amorces (A1, A2, A3, A4, A7, A13) donnant les profils d'amplification les plus simples (Fig.8), et le troisième une autre combinaison B de 6 amorces (A1, A2, A7, A10, A17, A19) parmi lesquelles les amorces A7, A10, A17 et A19 ont été utilisées par Tibayrenc *et al.*(1993) sur *T. cruzi*.

Le dendrogramme obtenu à partir de nos données isoenzymatiques et ceux obtenus à partir des données RAPD présentent de fortes similitudes (Fig.6 et 7) :

- Ils rassemblent tous les stocks du groupe I défini par Ochman *et al.*. (1984) en un ensemble homogène. Ce groupe est particulièrement bien identifié sur le dendrogramme élaboré à partir des 12 amorces.
- 5 génotypes répétés (R52, R54, R57, R58 et R59) apparaissent sur le dendrogramme de la combinaison A. Ces stocks sont également groupés sur les deux autres dendrogrammes élaborés à partir des données RAPD. Ces mêmes stocks (R54, R57, R58) présentent le même profil électrophorétique (de même que les stocks R13-14, R36-R40) et sont associés sur le dendrogramme construit avec les données d'Ochman *et al.* (1984). Ces résultats sont conformes à ceux d'Ochman *et al.*(1984) qui les plaçaient dans le groupe III (Annexe 5).
- Les autres stocks forment un ensemble plus hétérogène et désorganisé.

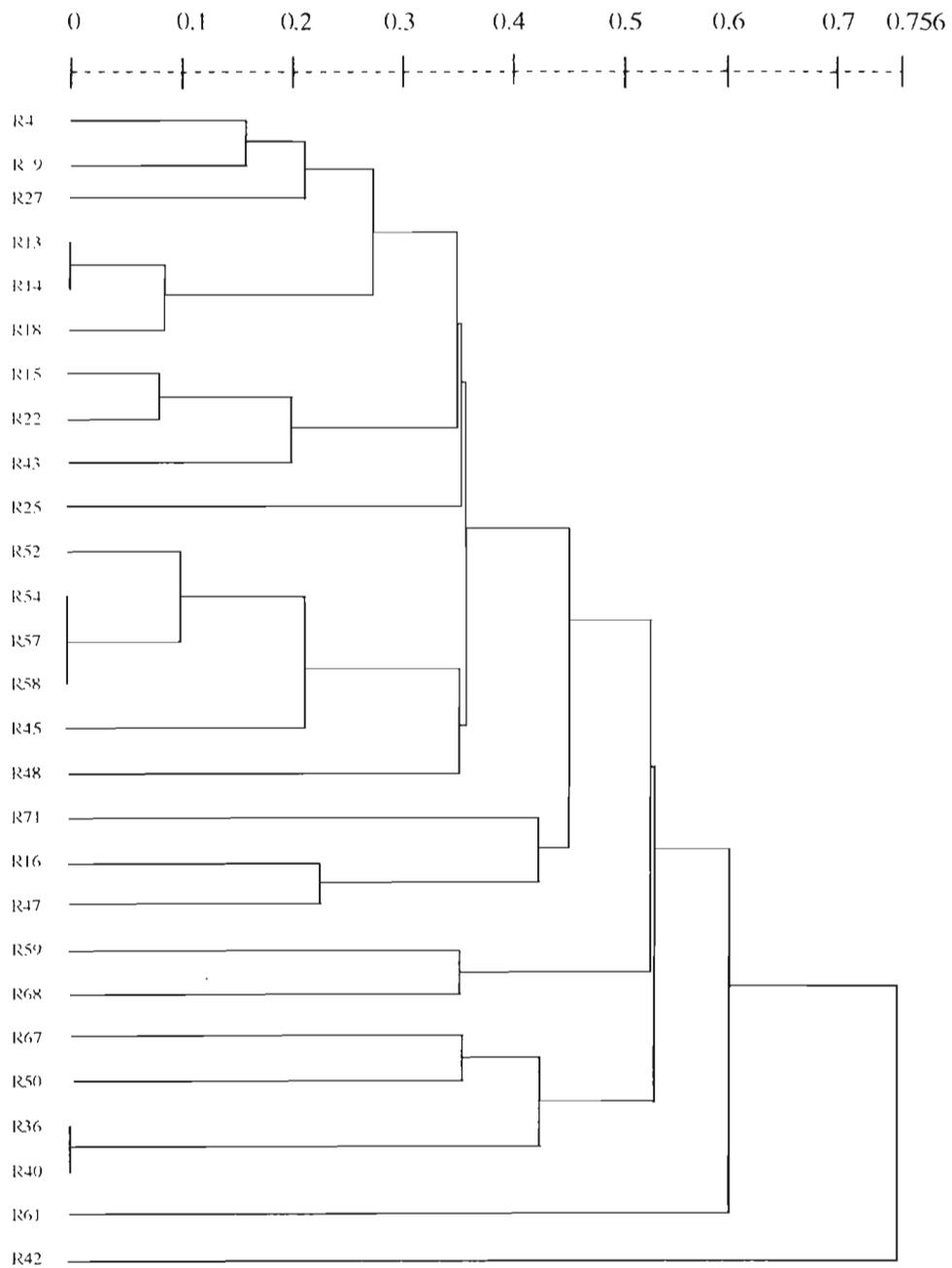


Fig.6 : Dendrogramme UPGMA élaboré à partir des distances de Jaccard calculées avec les données isoenzymatiques de *E. coli*.

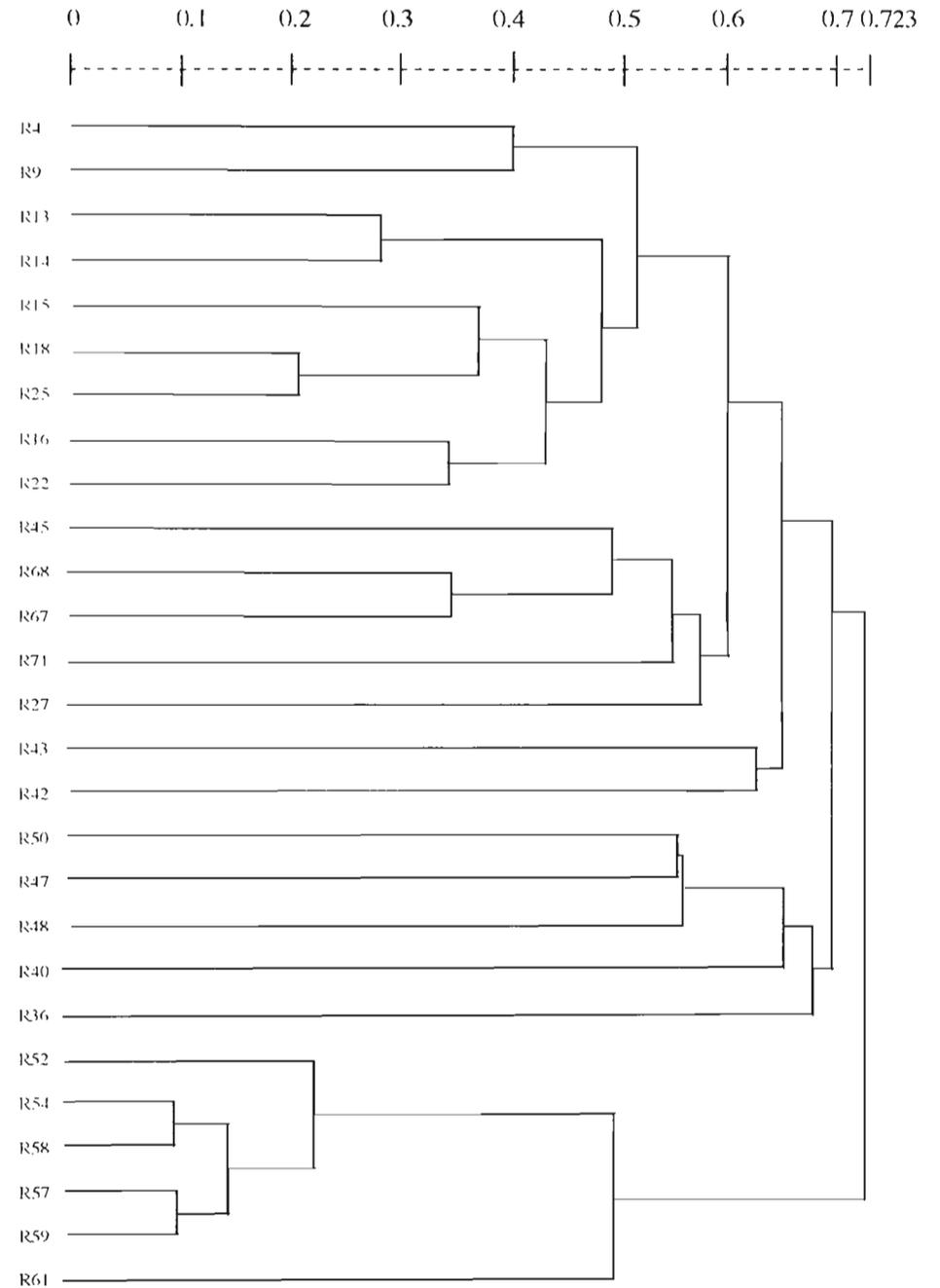


Fig.7 : Dendrogramme UPGMA élaboré à partir des distances de Jaccard calculées avec les données RAPI (12 amorces) de *E. coli*.

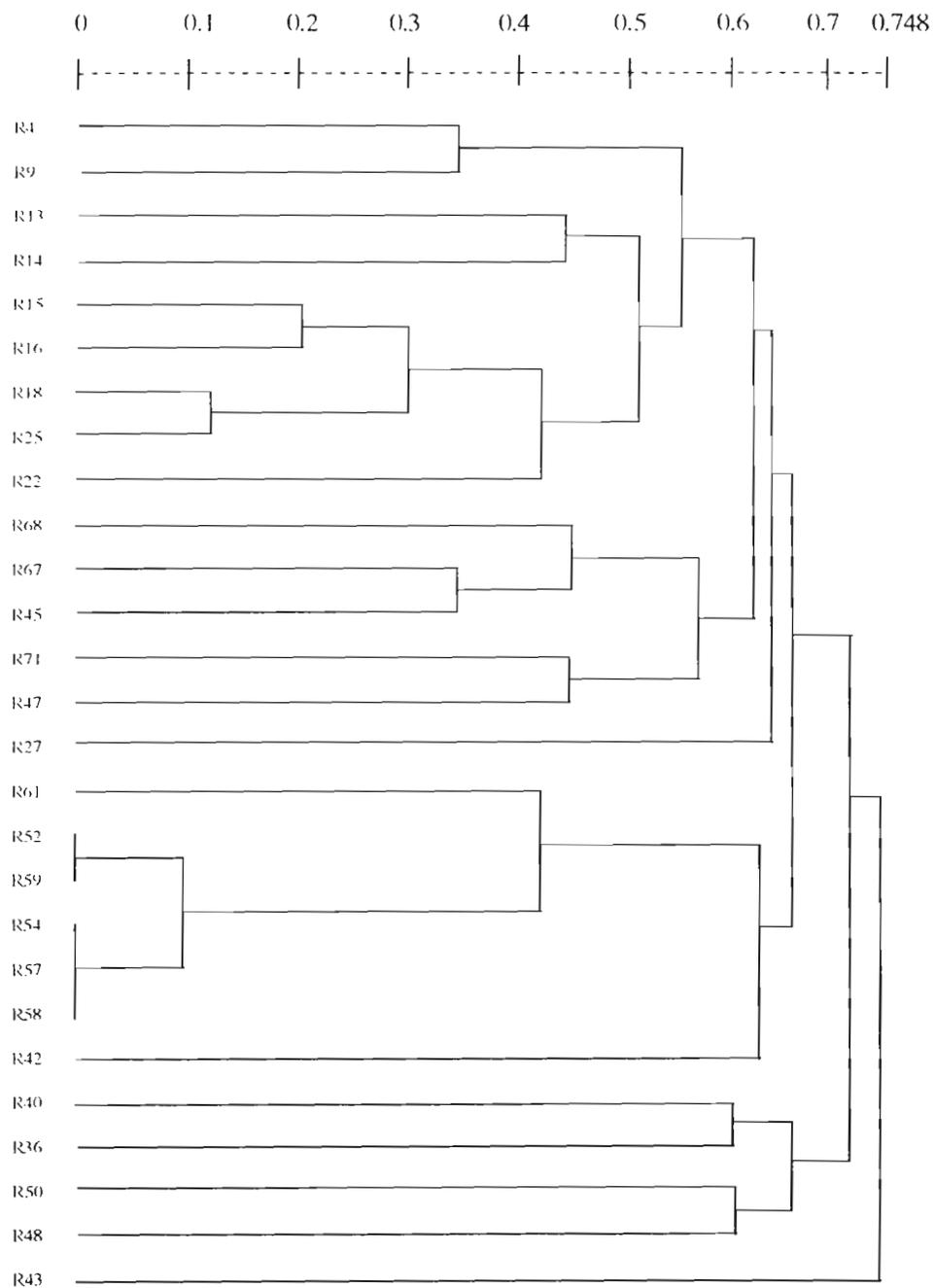


Fig.8 : Dendrogramme UPGMA élaboré à partir des distances de Jaccard calculées avec les données RAPD (combinaison d'amorces A) de *E. coli*

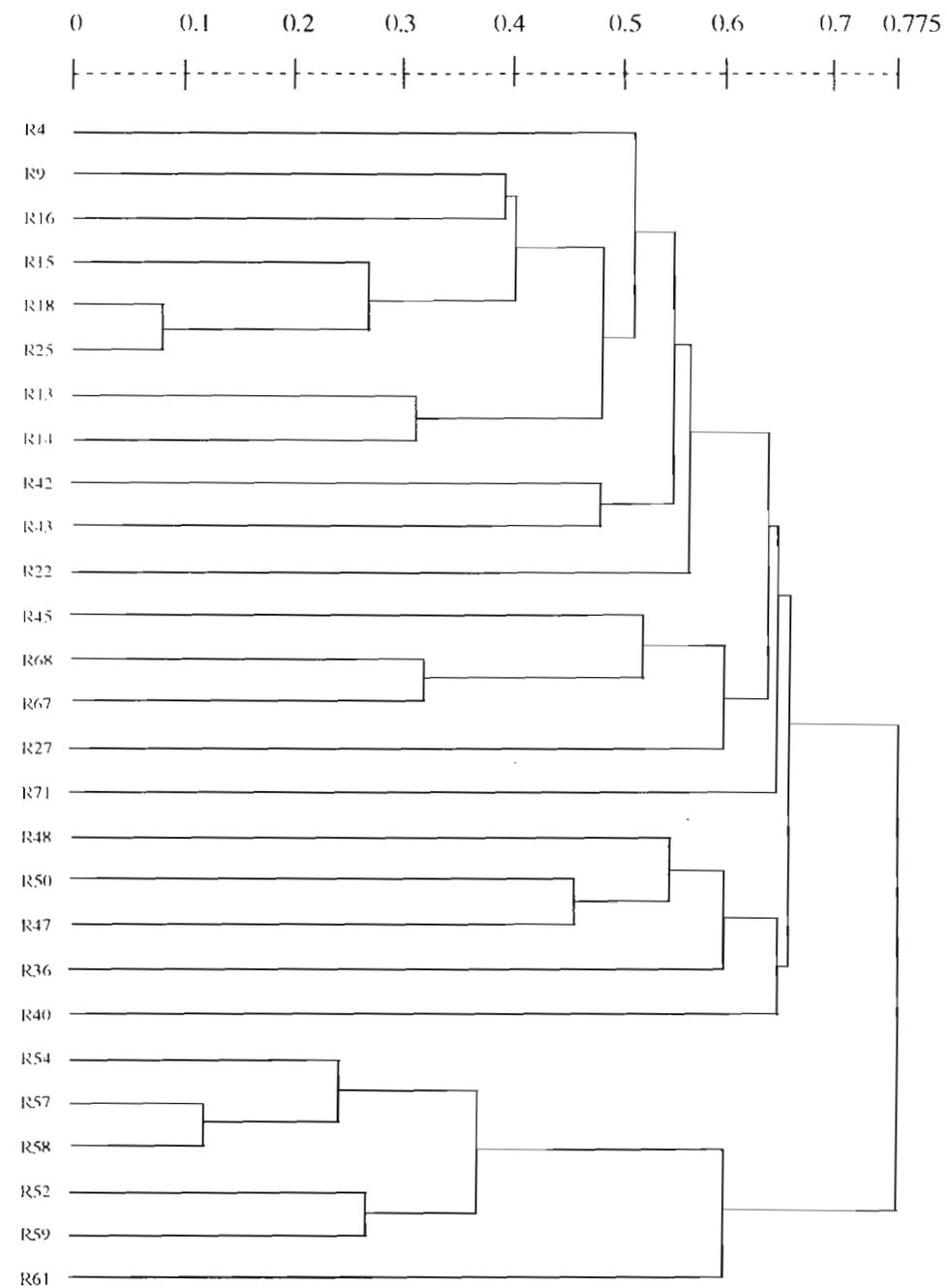


Fig.9 : Dendrogramme UPGMA élaboré à partir des distances de Jaccard calculées avec les données RAPD (combinaison d'amorces B) de *E. coli*.

5.1.5 Les tests statistiques

Ces tests prennent tous pour hypothèse nulle une situation de panmixie. Ils sont tous apparus non significatifs avec nos données isoenzymatiques :

-le test d1 avec une probabilité $p = 7.12 \times 10^{-2}$ de trouver un certain génotype répété trois fois ou plus (Annexe 6).

-le test d2 avec une probabilité $p = 0.21$ de trouver un génotype quelconque répété trois fois ou plus (Annexe 6).

-le test e avec une probabilité $p = 0.07$ de trouver 23 génotypes différents (Annexe 7).

-le test f avec une probabilité $p = 0.053$ d'observer un déséquilibre de liaison (Annexe 7).

Par ailleurs, ils apparaissent tous significatifs lorsqu'on utilise les données isoenzymatiques d'Ochman *et al.* (1984).

En ce qui concerne les RAPD, les tests d1, d2 et e ne sont pas utilisables avec les 12 amorces et la combinaison B étant donné qu'aucun génotype n'est répété. Le test d1 est par contre significatif avec la combinaison A avec une probabilité $p = 2.15 \times 10^{-5}$ et $p = 1.5 \times 10^{-7}$ de trouver un génotype particulier répété deux fois ou plus et une probabilité $p = 1.5 \times 10^{-5}$ de trouver un génotype répété trois fois ou plus. Les tests d2 et e sont aussi significatifs avec cette combinaison A avec des probabilités $p = 0.0031$ de trouver un génotype quelconque répété 2 fois ou plus et 26 génotypes différents au moins respectivement. Le test f est significatif avec une probabilité $p < 10^{-4}$ de trouver un déséquilibre de liaison avec l'ensemble des 12 amorces (Annexe 8) et la combinaison A, et non significatif ($p = 0.15$) avec la combinaison B. Le test g révèle une corrélation ($r = 0.4$, $P < 10^{-4}$) entre les deux matrices de distance fournies par les isoenzymes et les RAPD aussi bien avec nos données enzymatiques qu'avec celles d'Ochman *et al.*, ce qui indique un déséquilibre de liaison. Appliqué à l'intérieur des deux subdivisions révélées sur les dendrogrammes, il devient non significatif avec le groupe 1 ($r = -0.08$, $p > 0,05$). Cet indice reste néanmoins significatif avec l'autre groupe (Annexe 8).

Le test f a aussi été appliqué en l'absence de génotypes répétés avec nos données enzymatiques, celles d'Ochman *et al.* (1984), et la combinaison A d'amorces. Il apparaît significatif uniquement avec les données isoenzymatiques d'Ochman *et al.*

5.2 *Trypanosoma cruzi*

5.2.1 Distances et diversité génétiques

Les indices génétiques calculés à partir des isoenzymes ($D = 0.63$, $H = 0.59$) et RAPD ($D = 0.73$, $H = 0.77$) de *T. cruzi* de Tibayrenc *et al.* (1993) sont comparables à celles obtenus avec *E. coli*. Nous avons pu constater que les différences de migration entre électromorphes successifs étaient plus grandes (de l'ordre de 5 mm) chez *T. cruzi* que chez *E. coli*. (Fig. 5).

Le nombre moyen de génotypes obtenus avec les données RAPD (= 8.83) est plus élevé que celui obtenu avec les isoenzymes (= 5). Les RAPD révèlent toujours plus de variabilité que les isoenzymes.

Les diversités génotypiques obtenues avec les deux séries de données sont comparables, et égales à 0.7 et 0.75 respectivement : 17 génotypes différents ont été retrouvés avec les isoenzymes, et 18 avec les RAPD.

5.2.2 Les dendrogrammes

Les deux dendrogrammes UPGMA obtenus à partir des deux séries de données présentent des topologies similaires. Il en ressort une subdivision en deux groupes : l'un rassemble les clones 19-20 et apparentés, tandis que l'autre rassemble les clones 39 et apparentés (Annexe 11). Cette subdivision apparaît très robuste si l'on regarde les grandes distances génétiques séparant les deux groupes.

5.2.3 Les tests statistiques

Ils apparaissent tous hautement significatifs. Le test g nous donne une très forte corrélation entre les deux matrices de distances ($r = 0,89$ avec $p < 10^{-4}$). Lorsqu'il est appliqué à l'intérieur des grandes subdivisions, révélées par les dendrogrammes, il reste encore hautement significatif, ce qui indique un fort déséquilibre de liaison (Annexes 6, 7 et 8).

Le test f appliqué sans les génotypes répétés reste significatif avec les deux séries de données (RAPD et isoenzymes) (Annexe 9).

V DISCUSSION

Notre étude a jeté les bases d'une comparaison de la structure génétique des populations de deux microorganismes très différents, *E. coli* (une bactérie) et *T. cruzi* (un protozoaire parasite). Pour cela, nous avons étudié 27 souches de *E. coli* sur la base de 10 systèmes enzymatiques et 12 amorces RAPD. Les résultats ont été interprétés en termes de génétique des populations avec des tests statistiques reposant sur l'analyse du déséquilibre de liaison. Cette analyse de génétique des populations a été complétée par une étude de la diversité génétique générale de cette bactérie, ainsi que par une analyse de sa diversité phylogénétique intraspécifique. Les mêmes analyses ont été conduites chez *T. cruzi*, sur des données préalablement publiées par Tibayrenc *et al.*(1993)

E. coli

Les dendrogrammes élaborés à partir de nos données isoenzymatiques et RAPD sur les 27 souches sélectionnées dans notre échantillon, révèlent l'existence d'un groupe homogène comprenant 9 stocks. Ces stocks appartiennent tous au groupe I d'Ochman *et al.* (1993), qui ont défini 3 groupes, et au groupe A de Desjardins *et al.* (1995), qui ont défini 4 groupes. Nos données confirment donc l'existence de ce groupe comme unité bien délimitée au sein de *E. coli*.

Un second groupe apparaît dans nos dendrogrammes, visible plus nettement avec les données issues des 12 amorces RAPD (Fig.7) et confirmé par un bootstrap avec un pourcentage de 92 %. Les stocks de ce groupe appartiennent tous au groupe III d'Ochman *et al.*(1984) et au groupe B2 de Desjardins *et al.* (1995). Nos données recoupent donc bien les résultats des auteurs précédemment cités.

Les autres stocks forment un ensemble hétérogène sans groupe bien défini.

Nous avons utilisé différents tests statistiques prenant pour hypothèse nulle une situation de panmixie au sein des populations naturelles de *E. coli*. Nos résultats montrent plusieurs indices d'un fort déséquilibre de liaison, ce qui est en faveur d'une structure clonale au sein des populations de cette bactérie, et corrobore les hypothèses proposées par d'autres auteurs (Selander et Levin 1980, Orskov et Orskov 1983).

- Les tests sont significatifs avec l'ensemble de nos données RAPD, à l'exception du test f (déséquilibre de liaison) avec la combinaison d'amorces B.

- Avec nos données isoenzymatiques, les tests se révèlent non significatifs, mais significatifs avec les données d'Ochman *et al.*(1984). Par ailleurs, nous observons une forte corrélation entre distances génétiques générées par RAPD d'une part, et par isoenzymes (nos données et celles d'Ochman) d'autre part. Statistiquement ceci indique que les stocks présentant un fort degré de ressemblance par les isoenzymes, le montreront également de façon très

apparente par les RAPD, et vice-versa. Il s'agit là d'un cas, très évocateur, de déséquilibre de liaison.

Nos tests indiquent donc un déséquilibre de liaison au sein de l'ensemble de notre échantillon constitué de 27 souches. Ceci est un argument indirect classique pour une structure clonale et vient en confirmation des hypothèses précédemment proposées par d'autres auteurs (Selander et Levin 1980).

L'échantillonnage utilisé ici est très allopatrique, et on pourrait penser que le déséquilibre de liaison est imputable à la distance géographique, qui par effet de dérive génétique, entraîne des différences de fréquence génétique d'une localité à une autre. Cependant Tibayrenc *et al.* (1991) ont montré que lorsque le déséquilibre de liaison est imputable à la distance géographique, les génotypes surreprésentés ont une tendance proportionnelle à être confinés dans des aires géographiques restreintes. Ce n'est pas le cas ici. Si nous considérons nos données enzymatiques, les stocks R36 et R40 (qui présentent le même génotype) viennent respectivement de Suède et des USA. Les stocks R54, R57, et R58 viennent soit de l'Iowa, soit de Washington. Si nous prenons les données d'Ochman *et al.* (1984), les stocks R1, R2 et R3 viennent de l'Iowa, de Suède, et du Massachussets, les stocks R8, R9, R10 et R11 viennent de l'Iowa, de Suède, de New-York et de Suède, etc.

Ce déséquilibre de liaison persiste sur un échantillon plus restreint comprenant tous les stocks n'appartenant pas au groupe I.

Par contre, le test g apparaît non significatif lorsqu'il est effectué au sein de la subdivision I précédemment définie. Ce résultat peut être imputé à deux facteurs, qui diminuent proportionnellement la puissance du test (Tibayrenc 1995):(i) un trop faible effectif de ce groupe (9 stocks),(ii) une diminution de la variabilité génétique par rapport à l'échantillon total.

Le test f apparaît aussi non significatif lorsque les génotypes répétés sont retirés de l'échantillonnage. Dans ce cas aussi, les mêmes raisons peuvent être invoquées.

Il est impossible actuellement de savoir si la non significativité des tests est due à la diminution de leur puissance par ces deux facteurs, ou à l'existence de flux géniques au sein de ce groupe I. Ces deux hypothèses devront être départagées par un échantillonnage plus important.

La structure clonale de la population de la bactérie *E. coli* est largement acceptée, mais l'importance et l'impact de la recombinaison à l'échelle évolutive reste au centre des débats.

De nombreux arguments viennent en effet appuyer la théorie de clonalité des populations de *Escherichia coli* :

- des coefficients de déséquilibre de liaison proches du maximum théorique même quand les échantillons ont été prélevés dans des aires restreintes (Caugant *et al.* 1981, Souza *et al.* 1992).
- des souches présentant le même profil électrophorétique peuvent être retrouvées dans des hôtes distants dans l'espace et dans le temps.

- des concordances entre des dendrogrammes générés par électrophorèses enzymatiques et des arbres phylogénétiques basés sur des séquences nucléotidiques (Milkman et Bridges 1990, Nelson et Selander 1992, cf test g de Tibayrenc *et al.* 1990).

- un faible taux de transfert horizontal des gènes.

En ce qui concerne ce dernier point, Milkman et Bridges (1990) ont estimé que la fréquence de recombinaison, de l'ordre de 5×10^{-2} transferts par nucléotide et par génération, n'était pas suffisante pour perturber la structure clonale des populations de bactéries. Les mutations ponctuelles étant 5 fois plus fréquentes que la recombinaison, la divergence clonale à partir d'un ancêtre commun serait due essentiellement au processus mutationnel.

Maynard Smith (1991) a souligné, chez les bactéries, l'existence de gènes mosaïques dont la structure résulterait d'échanges de fragments de gènes entre espèces proches. Plusieurs exemples ont été décrits chez des espèces de bactéries aptes à la transformation naturellement telles que *Neisseria* et *Haemophilus*, soumises à de fortes pressions de sélection, pour lesquelles l'acquisition d'un gène nouveau peut apporter un avantage sélectif. Ce qui n'a pas empêché certains clones de *N. meningitidis* et de *H. influenzae* de se répandre car les effets de la transformation sont locaux et n'altèrent qu'occasionnellement les types électrophorétiques. Le transfert horizontal de gènes a aussi été invoqué pour expliquer l'existence de gènes mosaïques chez des bactéries non transformantes telles que *E. coli* par Stolfus *et al.* (1988). Mais ces échanges ponctuels et limités à une centaine de paires de bases, sont habituellement considérés comme trop rares pour perturber la structure clonale de ces bactéries.

Cependant, Guttman et Dykhuizen (1994) ont réalisé une étude phylogénétique basée sur des généalogies de gènes. Le séquençage de 4 gènes différents (*sppA*, *gapA*, *pabB* et *zwf*) situés dans la même région du chromosome a montré que trois événements de recombinaison sont survenus dans l'histoire évolutive de 5 souches du groupe I depuis leur divergence à partir d'un ancêtre commun. Le taux de recombinaison a été estimé à 5×10^{-9} changements par nucléotides par génération, soit 50 fois plus que le taux de mutations. La longueur du fragment recombinant est estimé à 45 kb. La recombinaison serait la seule responsable de la divergence du groupe I.

L'hypothèse de Guttman et Dykhuizen (1994) est difficile à accepter au vu des résultats présentés par d'autres auteurs (données de la littérature et résultats obtenus dans le présent travail). Si les événements de recombinaison étaient aussi fréquents que ne le suggère le travail de ces auteurs, l'espèce *Escherichia coli* devrait tendre à constituer un pool génique unique (comme cela semble être le cas pour *Neisseria gonorrhoeae*), et ne devrait pas montrer la forte structuration que suggèrent les taux importants de déséquilibre de liaison observés. Mais cette hypothèse apparaît néanmoins compatible avec les résultats de certains tests statistiques qui suggèrent l'existence de flux géniques au sein du groupe I.

T. cruzi

Les tests d1, d2, e et f sont hautement significatifs pour les deux séries de marqueurs. Les dendrogrammes mettent en évidence 2 grands sous-ensembles bien distincts au sein desquels persiste le déséquilibre de liaison.

De précédents travaux ont montré que la recombinaison est restreinte ou totalement absente dans les populations naturelles de *T. cruzi*. Certains zymodèmes montrent une hétérozygotie fixée à plusieurs locus pendant de longues périodes de temps et de vastes aires géographiques (Tibayrenc et Ayala 1988). Ceci est incompatible avec la ségrégation méiotique. Le zymodème 39, par exemple, est hétérozygote aux locus de la GPI, la PGM et la 6PGD : il reste inchangé sur de vastes aires géographiques incluant le Chili, de nombreuses localités de Bolivie et du Sud du Brésil, et a été isolé de nombreuses fois entre 1977 et 1984. De nombreux génotypes sont aussi retrouvés en sympatrie dans la même maison, le même hôte et dans le même insecte vecteur, ce qui laisse toute opportunité pour d'éventuels échanges génétiques. 10 % des *Triatoma infestans* de Bolivie, par exemple, renferment des mélanges de stocks de deux zymodèmes différents (Tibayrenc *et al.* 1986). Mais aucun hybride résultant de la recombinaison de ces zymodèmes n'a encore été retrouvé ni en Bolivie, ni ailleurs.

Approche comparative entre *E. coli* et *T. cruzi*

L'objectif général de cette étude était de comparer l'impact de la recombinaison sur la diversité et l'évolution de deux microorganismes. Nous avons utilisé les mêmes techniques moléculaires (isoenzymes et RAPD) et les mêmes tests avec *E. coli* et *T. cruzi*.

A l'heure actuelle, aucun indice de recombinaison génétique n'a été relevé chez *T. cruzi*, contrairement à la découverte de gènes mosaïques chez *E. coli*. Dans cette optique, il serait intéressant de réaliser des généalogies de plusieurs gènes chez *T. cruzi* afin de tester leur congruence.

Nos tests ont indiqué des indices de déséquilibre de liaison plus forts et plus constants chez *T. cruzi*, suggérant que les populations de ce parasite présentent une structure davantage clonale que celles d'*E. coli*. Cependant la comparaison reste préliminaire.

Les RAPD ont révélé au sein des deux espèces étudiées un polymorphisme génétique plus important que les électrophorèses d'isoenzymes et caractérisé par des distances génétiques plus grandes et un nombre moyen de génotypes plus grand. Les différents indices génétiques calculés sont comparables chez les deux espèces de microorganismes considérés, exceptés : le nombre moyen de génotypes et la diversité génotypique qui sont supérieurs chez *E. coli*. Ceci

indiquerait une plus grande variabilité génétique au sein des isolats de *E. coli* et un déséquilibre de liaison moins fort, conformément aux résultats des tests statistiques.

Par ailleurs, l'analyse phénétique n'a pas permis de mettre en évidence les trois groupes subspécifiques définis par Ochman *et al.* (1984) au sein des stocks de référence de *E. coli*. Seuls deux groupes correspondant au groupe I et III d'Ochman *et al.* (1984) semblent se détacher. L'analyse phénétique a néanmoins révélé l'existence d'une répartition en deux groupes des stocks de *T. cruzi*.

Un autre élément a pu être relevé dans cette étude : les différences de migration entre électromorphes successifs sont nettement plus grandes chez *T. cruzi* (de l'ordre de 5 mm) que chez *E. coli* (1 mm). Ceci résulterait d'une accumulation plus importante de mutations au sein de cette espèce.

Ces différentes constatations suggèrent une divergence phylogénétique intraspécifique plus grande chez *T. cruzi*, bien que la distance de Jaccard ne le mette pas en évidence.

Les RAPD ont révélé l'existence de bandes partagées par l'ensemble des stocks chez *E. coli* (bandes générées par les amorces A1, A2 et A16) ou uniquement par certains groupes de stocks chez les deux espèces (ex: l'amorce A10). Ces bandes communes constituent des caractères synapomorphes que l'on retrouverait chez le plus récent ancêtre commun. L'amorce A10, par exemple, montre trois bandes caractéristiques chacune d'un des trois groupes subspécifiques de l'espèce *E. coli* définis par Ochman *et al.*. Ces différentes amorces pourraient être utilisées comme amorces-diagnostic de l'espèce. De même chez *T. cruzi* (Tibayrenc *et al.* 1993), de nombreuses amorces-diagnostic (une trentaine environ dont l'amorce A8) donnant des bandes amplifiés chez tous les stocks ou sous groupes de stocks mais absentes chez des espèces voisines telles que *T. cruzi marinkellei* ont été découvertes.

VI CONCLUSION

L'étude présentée ici doit être considérée comme préliminaire, et devra obligatoirement être complétée d'une nouvelle ligne de résultats après une nouvelle phase de mise au point technique, portant en particulier sur l'électrophorèse d'isoenzymes sur acétate de cellulose, employée pour la première fois chez *E. coli*.

(a) Mise au point de marqueurs génétiques

L'électrophorèse d'isoenzymes sur acétate de cellulose a donné des résultats déjà exploitables chez *E. coli*, et présente par ailleurs les mêmes avantages que pour les protozoaires parasites (Ben Abderrazak *et al.* 1993) : simplicité d'emploi et rapidité. Une phase additionnelle de mise au point s'avère cependant nécessaire avant que l'outil ne puisse être considéré comme opérationnel : un minimum de 15 locus apparaît requis pour l'analyse de génétique des populations et de phylogénétique.

Le RAPD donne des résultats satisfaisants et exploitables : l'information phylogénétique semble robuste, comme le montre la corrélation significative entre distances génétiques isoenzymatiques et RAPD, résultat préalablement observé pour *T. cruzi* (Tibayrenc *et al.* 1993). Le RAPD présente le grand avantage de reposer sur une technique standard, utilisable quel que soit l'organisme à l'étude, alors que les isoenzymes nécessitent à chaque fois une mise au point spécifique.

(b) Structure des populations

Aussi bien pour les données publiées par Ochman *et al.* (1984) que pour les résultats que nous avons produits, les tests de déséquilibre de liaison développés à l'origine pour les protozoaires parasites (Tibayrenc *et al.* 1990) ont montré chez *E. coli* des résultats fortement positifs, ce qui corrobore l'hypothèse d'une structure clonale présentée par Selander et Levin (1980). La présence de génotypes répétés ubiquistes suggère que le déséquilibre de liaison chez *E. coli* n'est pas imputable à la seule distance géographique. Certains tests de déséquilibre de liaison (tests pratiqués au sein de chaque groupe, ou en éliminant les génotypes répétés) se sont révélés négatifs. Ce résultat, qui n'a jamais été observé chez *T. cruzi*, pourrait suggérer deux hypothèses:

- (i) tests devenant non significatifs quand ils sont pratiqués au sein d'un groupe au lieu de l'espèce entière : le groupe représente une sexualité cryptique (Maynard Smith *et al.* 1993).
- (ii) tests devenant non significatifs quand ils sont pratiqués après avoir éliminé les génotypes répétés : la population à l'étude présente une structure épidémique (propagation de clones éphémères au sein d'une espèce fondamentalement sexuée)

Cependant, à ce stade de notre étude, ces tests, basés du reste sur une approche qui pourrait présenter certains biais (Tibayrenc 1995 Adv. Parasitol.), doivent être interprétés avec

prudence : impossibilité de rejeter l'hypothèse nulle (ici: absence de déséquilibre de liaison) n'est pas confirmation de cette hypothèse, et peut-être dû simplement à un manque de puissance du test.

Il n'en demeure pas moins que la structuration des populations d'*E.coli* semble moins forte que celle des populations de *T. cruzi*. Un tel résultat était impossible à affirmer sans l'approche comparative directe développée ici.

(c) Analyse phylogénétique

Faute de temps, nous n'avons pas pu mettre en œuvre les techniques classiques d'analyse phylogénétique, et nous avons dû nous contenter d'élaborer des dendrogrammes UPGMA. Cependant, la corrélation très significative mise en évidence entre distances isoenzymatiques et distances RAPD suggère que les deux groupes révélés par cette analyse au sein des stocks de *T. cruzi* représentent des lignées évolutives distinctes.

Les groupes que nous observons chez *E. coli* recourent dans une certaine mesure ceux mis en évidence par Ochman *et al.* (1984) bien que contrairement à ces auteurs, nous n'en observons que deux. Ici cependant, notre technique isoenzymatique devra être perfectionnée avant que nous ne puissions considérer nos analyses phylogénétiques comme véritablement exploitables.

Les distances de Jaccard ne prennent en compte que le paramètre présence-absence des bandes, sans tenir compte de la distance de migration des électromorphes. Nous avons constamment observé que les électromorphes d'*E. coli* sont très proches les uns des autres, contrairement à ce qui est observé chez *T. cruzi*. Or il a été suggéré que la plus ou moins grande différence de migration d'électromorphes reflétait des divergences phylogénétiques plus ou moins grandes (Briscoe *et al.* 1987). Dans le laboratoire d'accueil, une distance génétique basée sur ce principe a été utilisée pour étudier la divergence phylogénétique au sein du genre *Phytomonas* (Guerrini 1989). Faute de temps, nous n'avons pu mettre en œuvre cette approche, qui aurait très probablement indiqué des distances génétiques nettement plus faibles au sein d'*E. coli* que chez *T. cruzi*.

La présente étude a permis de jeter les bases de l'approche comparative entre microorganismes qui constitue le fondement à venir du laboratoire d'accueil. D'une manière générale, cette approche comparative aussi directe que possible a permis de montrer, aussi bien pour l'analyse de génétique des populations que pour celle de phylogénétique, que les similitudes entre les profils évolutifs d'*E. coli* et *T. cruzi* n'étaient pas aussi poussées que ne pouvait le laisser penser l'analyse des seules données de la littérature. Notre travail de thèse visera à compléter ce travail préliminaire, et à l'enrichir par la prise en compte d'autres espèces bactériennes.

Références bibliographiques

- Anilionis A. et Riley M. (1980). Conservation and variation of nucleotide sequences within bacterial genomes: *Escherichia coli* strains. *J. Bacteriol.* **143**, 355-365.
- Ben Abderrazak S., Guerrini F., Mathieu-Daudé F., Truc P., Neubauer K., Lewicka K., Barnabé C. et Tibayrenc M. (1993). Isoenzyme electrophoresis for parasite characterization. *Methods in Molecular Biology* **21**, 361-382.
- Briscoe D. A., Cooley A. A., Bernstein R. L., Mac Key G. M. et Williams K. L. (1987). Genetic diversity in cellular slime molds: allozyme electrophoresis and a monoclonal antibody reveal cryptic species among *Dictyostelium discoideum* strains. *Genetics* **117**, 213-220.
- Caugant D. A., Levin B. R. et Selander R. K. (1981). *Genetics* **98**, 467-490.
- Cavé H., Bingen E., Elion J. et Denamur E. (1994). Differentiation of *Escherichia coli* strains using randomly amplified polymorphic DNA analysis. *Res. Microbiol.* **145**, 141-150.
- Desjardins P., Picard B., Kaltenböck B., Elion J. et Denamur E. (1995). Sex in *Escherichia coli* does not disrupt the clonal structure of the population: evidence from RAPD and *rrn* RFLP data. *Journal of Molecular Evolution* sous presse.
- Eveland W. C., Oliver W. J. et Neel J. V. (1971). Characteristics of *Escherichia coli* in Yanomama, a primitive indian tribe of South America. *Infection and Immunity* **4**, 753-756.
- Guerrini F. (1989). Variabilité génétique et taxonomie du genre *Phytomonas*, Trypanosomatidae, parasites de plantes utiles. *Diplôme d'Etudes Approfondies de Parasitologie* Université de Montpellier II Sciences et Techniques du Languedoc.
- Guttman D. S. Dykhuizen, D.E. (1994). Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science* **266**, 1380-83.
- Harshman L. et Riley M. (1980). Conservation and variation of nucleotide sequences in *Escherichia coli* strains isolated from nature. *J. Bacteriol.* **144**, 560-568.
- Hartl D. L. et Dykhuizen D. E. (1984). The population genetics of *Escherichia coli*. *Annual Review of Genetics* **18**, 31-68.
- Lanar D. F., Levy L. S. et Manning J. E. (1981). Complexity and content of the DNA in *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* **3**, 327-341.
- Mantel N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209-220.
- Maynard Smith J., Dowson C. G. et Spratt B. G. (1991). Localized sex in bacteria. *Nature* **349**, 29-31.

- Maynard Smith J. Smith, N.H., O'Rourke, M., Spratt, B.G. (1993). How clonal are bacteria? *Proc. Natl. Acad. Sci. USA* **90**, 4384-88.
- Milkman R. (1973). Electrophoretic variation in *Escherichia coli* strains from natural sources. *Science* **182**, 1024-26.
- Milkman R. et Crawford I. P. (1983). Clustered third-base substitutions among wild strains of *Escherichia coli*. *Science* **221**, 378-380.
- Milkman R. et McKane Bridges M. (1990). Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. *Genetics* **126**, 505-517.
- Nelson K. et Selander R. K. (1992). *J. Bacteriol.* **174**, 6886-95.
- Ochman H., Whittam T., Caugant D. A. et Selander R. K. (1983). Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. *Journal of General Microbiology* **129**, 2715-2726.
- Ochman H. et Selander R. K. (1984). Standard reference strains of *Escherichia coli* from natural populations. *Journal of Bacteriology* **157**, 690-693.
- Orskov F. et Orskov I. (1983). Summary of a workshop on the clone concept in the epidemiology, taxonomy, and evolution of the Enterobacteriaceae and other Bacteria. *J. Infect. diseases* **148**, 346-357.
- PNUD, Mondiale B. et TDR O. (1993). Eleventh Programme Report. OMS, Geneva, Pujol C., Reynes J., Renaud F., Mallié M. et Bastide J. M. (1993). Analyse génétique de souches de *Candida albicans* par électrophorèse des isoenzymes. *J. Mycol. Med.* 14-19.
- Selander R. K. et Levin B. R. (1980). Genetic diversity and structure in *Escherichia coli* populations. *Science* **210**, 245-247.
- Selander R. K., Korhonen T. K., Vaisanen-Rhen V., Williams P. H., Pattison P. E. et Caugant D. A. (1986). Genetic relationship and clonal structure of strains of *Escherichia coli* causing neonatal septicemia and meningitis. *Infection and Immunity* **52**, 213-222.
- Sneath P. H. A. et Sokal R. R. (1973). Numerical taxonomy. The principle and practice of numerical classification. Freeman, San Francisco. 537.
- Souza V., Nguyen T. T., Hudson R. R., Piñero D. et Lenski R. E. (1992). *Proc. Natl. Acad. Sci. USA* **89**.
- Tibayrenc M., Cariou M. L. et Salignac M. (1981). Interprétation génétique des zymogrammes de flagellés des genres *Trypanosoma* et *Leishmania*. *C. R. Acad. Sc. Paris* **292**, 623-625.

Tibayrenc M., Ward P., Moya A. et Ayala F. J. (1986). Natural populations of *Trypanosoma cruzi*, the agent of Chagas' disease, have a complex multiclonal structure. *Proc. Natl. Acad. Sci. USA* **83**, 115-119.

Tibayrenc M. Ayala, F.J. (1988). Isozyme variability in *Trypanosoma cruzi*, the agent of Chagas' disease: genetical, taxonomical, and epidemiological significance. *Evolution* **42**, 277-292.

Tibayrenc M. Kjellberg, F., Ayala, F.J. (1990). A clonal theory of parasitic protozoa: the population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. *Proc. Natl. Acad. Sci. USA* **87**, 2414-18.

Tibayrenc M. et Ayala F. J. (1991). Towards a population genetics of microorganisms: the clonal theory of parasitic protozoa. *Parasitology Today* **7**, 228-232.

Tibayrenc M. Neubauer, K., Barnabé, C., Guerrini, F., Skarecky, D., Ayala, F.J. (1993). Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. *Proc. Natl. Acad. Sci. USA* **90**, 1335-39.

Tibayrenc M. (1995). Population genetics of parasitic protozoa and other microorganisms. *Advances in Parasitology* **36**, 47-115.

Tibayrenc M. (1995). Population genetics and strain typing of microorganisms: how to detect departures from panmixia without individualizing alleles and loci. *C.R. Acad. Sci. Paris, Sciences de la vie* **318**, 135-139.

Welsh J. et McClelland M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* **18**, 7213-18.

Williams J. G. K., Kubelik A. R., Livak K. J., Rafalski J. A. et Tingey S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**, 6531-35.

ANNEXES

ANNEXE 0

Les différentes applications de la génétique des populations des microorganismes

Table 1 The applied and basic aspects of population genetics of microorganisms.

Main applications of population genetics and evolutionary studies in microbiology	Translation in terms of basic science
<p><i>Epidemiological tracking</i> Checking for the stability of microbial genotypes over space and time Short-term level: nosocomial epidemiology^a Long-term level: broad-scale epidemiology^b</p>	<p>Structure and dynamics of microbial populations Impact of genetic recombination on population structure; evolutionary role of sex</p>
<p><i>Taxonomy</i> Must be first based upon phylogeny^c Exploring the relationships between genetic diversity and the commonly accepted taxonomical nomenclature Looking for hidden genetic subdivisions within presently identified species</p>	<p>Molecular phylogeny; evolutionary role of sex</p>
<p><i>Studies downstream from genetics</i> Impact of genetic diversity and phylogenetic divergence on the relevant properties of microorganisms Virulence, resistance to drugs, immunological patterns, susceptibility to potential vaccines Vector and host specificity</p>	<p>Adaptative significance of microbial genetic diversity Vector/host/parasite co-evolution</p>

^a Time and space scales: days–months, hospital based.

^b Time and space scales: months–years, village based.

^c Time and space scales: millions of years, country- or continent-wide, up to the whole geographical range of the species.

ANNEXE 1

Noms et Codes d'identification des systèmes enzymatiques étudiés :

Nom	Abréviation	code international
Gycéraldéhyde-3-phosphate deshydrogénase	GAPD	E.C.1.2.1.12.
Glutamate oxaloacétate transaminase	GOT	E.C.2.6.1.1.
Glucose-6-phosphate déhydrogénase	G6PD	E.C.1.1.1.49.
Glucose -phosphate isomérase	GPI	E.C.5.3.1.9.
Malate déshydrogénase	MDH	E.C.1.1.1.37.
Mannose-phosphate isomérase	MPI	E.C.5.3.1.8
Nucléoside hydrolase (inosine)	NHI	E.C.2.4.2.-.
Nucléoside phosphorylase	NSP	E.C.2.4.2.1.
Phosphogluconate déshydrogénase	6PGD	E.C.1.1.1.44.
Superoxide dismutase	SOD	E.C.1.15.1.1.

Conditions électrophorétiques:

Enzyme	Tampon	Imbibition dilution	Migration dilution	Voltage	Temps
GAPD	A sel	0,1	1	100 V	20'
GOT	A sel	0,1	1	70 V	20'
G6PD	HR	0,5	1	160 V	20'
GPI	III Shaw	0,2	1	160 V	20'
MDH	III Shaw	0,3	1	120 V	20'
MPI	HR	0,5	1	160 V	20'
NHI	DR2X	1	1	200 V	15'
NSP	A sel	0,08	1	100 V	20'
6PGD	A sel	0,1	1	100 V	20'
SOD	SOD	1	1	160 V	20'

Tampons de migration et d'imbibition :

Tampon	Désignation	Composition
A sel	Selander A	Tris 0,68 M, Acide citrique monohydrate 0,01 M, pH 8.0
Dr2X	Richardson D x2	Tris 15 mM, EDTA 5mM, MgCl ₂ 10 mM, Acide borique 5,5 mM pH 7,8
HR	HR	Tris-Barbital-Sodium Barbital (HR Helena) pH 9
III Shaw	Shaw n° III	Tris 0,5 M, Acide borique 0,65 M, EDTA 0,016 M, pH 8.0
SOD	SOD	Tris 0,1 M, Acide borique 0,3 M, pH 7.6*

* nécessite l'addition de 10 % de Sucre

ANNEXE 2

Solutions de révélation :

Enzyme	Tampon de réaction	Solution de coloration
GAPD	2.5 ml Tris HCL 1 M pH 8, eau qsp 10 ml	fructose 1,6-diphosphate 12.5 mg, aldolase 0.7 UI, ac.arsenique 12.5 mg, ac. pyruvique 12.5 mg, NAD 4 mg, NBT 2.5 mg, PMS 3 mg
GOT	10 ml Tris HCL 0.1 M pH 8	L.ac.aspartique 25 mg, ac. α -Kétoglutarique 20 mg, pyridoxal 5-phosphate 3 mg, fast blue BB salt 20 mg
G6PD	5 ml Tris HCL 1 M pH 7, eau qsp 10 ml	D-glucose-6-phosphate 5mg, MgCl ₂ 50 mg, NADP 5mg, NBT 2.5 mg, PMS 3 mg
GPI	2.5 ml Tris HCL 1 M pH 7, eau qsp 10 ml	Fructose-6-phosphate 10 mg, glucose-6-phosphate déshydrogénase 5 UI, MgCl ₂ 50 mg, NADP 5 mg, NBT 2.5 mg, PMS 3mg
MDH	1 ml Tris HCL 1 M pH 8, eau qsp 10 ml	ac.malique 1 M pH 7 0.9 ml, NAD 4 mg, NBT 2.5 mg, PMS 3 mg.
MPI	2.5 ml Tris HCL 1 M pH 7, eau qsp 10 ml	mannose-6-phosphate 20 mg, glucose phosphate isomérase 10 UI, glucose 6 phosphate déshydrogénase 10 UI, Mgcl ₂ 50 mg, NADP 5 mg, NBT 2.5 mg, PMS 3 mg.
NHI	10 ml Tris HCL 0.1 M pH 8	inosine 10 mg, xanthine oxydase 0,6 UI, NBT 2.5 mg, PMS 3 mg.
NSP	10 ml tampon Phosphate 0.1 M	Xanthine oxydase 0.8 mg, NBT 2.5 mg, PMS 3 mg
6PGD	2.5 ml Tris HCL 1 M pH 8, eau qsp 10 ml	6-phosphogluconique 10 mg, MgCl ₂ 50 mg, NADP 5 mg, NBT 2.5 mg, PMS 3 mg
SOD	10 ml tampon phosphate 0.1 M pH 7.5	EDTA 7.5 mg, INT 3.5 mg. Après la migration, conserver à l'abri de la lumière et à température ambiante pendant 10 min, puis exposer à une lumière fluorescente pour la révélation des bandes.

Séquences nucléotidiques des amorces :

code	5'	to	3'
A1		CAGGCCCTTC	
A2		TGCCGAGCTG	
A3		AGTCAGCCAC	
A4		AATCGGGCTG	
A7		GAAACGGGTG	
A9		GGGTAACGCC	
A10		GTGATCGCAG	
A13		CAGCACCCAC	
A16		AGCCAGCGAA	
A17		GACCGCTTGT	
A19		CAAACGTCGG	
A20		GTTGCGATCC	

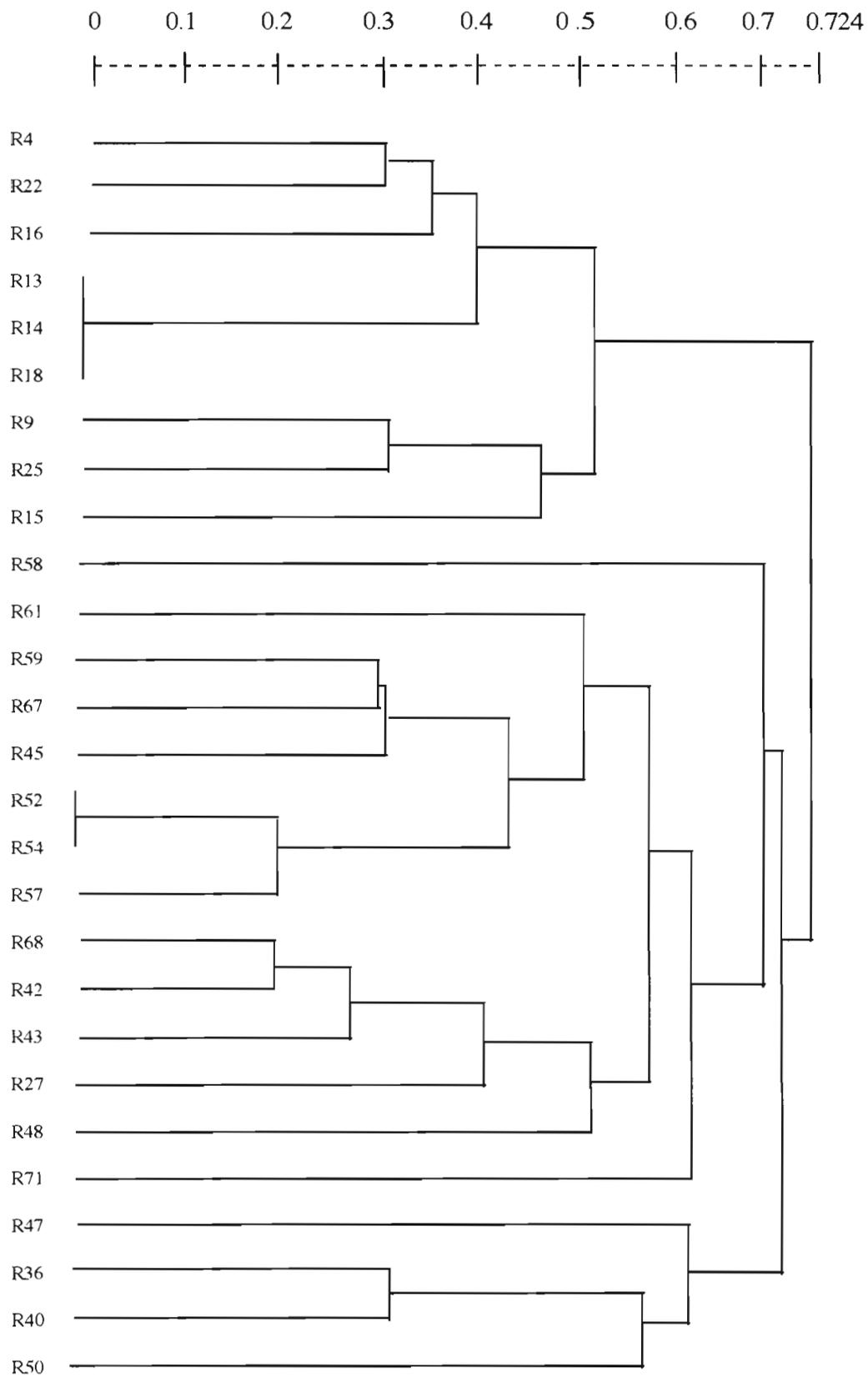
ANNEXE 3 :

Zymodèmes des 27 stocks de *E. coli* caractérisés sur la base de 10 systèmes enzymatiques

Souches	GOT	NSP	GPI	GAPD	SIXPG	SOD	MDH	G6PD	NHI	MPI
R4	1	1	2	1	3	1-4	1	2	2	1
R9	1	1	2	1	2	1-4	1	2	2	1
R13	1	1	2	1	2	1-4	1	1	2	1
R14	1	1	2	1	2	1-4	1	1	2	1
R15	1	1	2	1	2	1	1	2	2	4
R16	1	1	2	1	1	1-4	1	2	2	4
R18	1	1	2	1	2	1	1	1	2	1
R22	1	1	2	1	2	1-5	1	2	2	4
R25	1	1	2	2	2	1	1	2	2	1
R27	1	1	2	1	2	1-4	1	2	2	3
R36	1	1	4	1	1	1	3	2	2	3
R40	1	1	4	1	1	1	3	2	2	3
R42	1	1	3	2	2	3	2	2	3	3
R43	1	1	2	1	2	1	1	2	2	2
R45	1	1	4	1	2	1-4	1	2	2	5
R47	1	1	2	1	1	1	4	2	2	4
R48	1	1	3	1	2	1	1	2	2	3
R50	1	1	5	1	3	1	2	2	2	3
R52	1	1	4	1	2	1-4	1	2	2	1
R54	1	1	4	1	2	1	1	2	2	1
R57	1	1	4	1	2	1	1	2	2	1
R58	1	1	4	1	2	1	1	2	2	1
R59	1	1	4	1	2	1	2	2	2	2
R61	1	1	4	1	3	1	1	1	1	1
R67	1	1	4	1	3	1	1	2	2	3
R68	1	1	2	1	2	1	2	2	1	2
R71	1	1	1	1	1	1	1	2	2	1

ANNEXE 4 : Rapdèmes des 27 stocks de *E. coli* caractérisés avec 12 amorces.

souches	A1	A2	A3	A4	A7	A9	A10	A13	A16	A17	A19	A20
R4	2	4-6	2-9	9-11-15	10	1-4-9	12-13	4-5	4-5	3-6-10	11-12-14	6-8-12
R9	2	4-6	2-9	9-11	10-13	1-3-4-5-9	10	1-2-5	2-4-5	6-7-10	11-12-14	6-8
R13	2	6	5-9	9-15	12	1-3-4-5-9	1-10	1-5	2-4-5	2-5-6-8-10	11-12-14	12
R14	2	6	2-5-9	9-12-14-15	4	1-3-4-5-9-11	2-10-13	1-2-3-5	2-4-5	2-5-6-8-10	11-12-14	12
R15	2	6	2-5-8-9	9-15-19	8-10	3-4-5-9	8-10	1-2-3-5	4-5	5-10	11-12-14	5-6-9-11-12
R16	2	6	2-8-9-11	4-9-15-19	10	3-4-5-9	10	1-2-3-5	3-5	6-8-10	11-12	8-12
R18	2	5-6	2-8-9-11	9-15-19	10	4-5-9	10	3-5	2-4	5-10	11-12-14	8-12
R22	2	6	2-5-8-9	9-18	5-10	3-4-5-9	5-6	1-3-5	3-5	6-10	11-12	8-12
R25	2	5-6	2-9	9-15-19	10	3-4-5-9	10	3-5	4-5	5-6-10	11-12-14	8-12
R27	1-2	1-6	1-2-9	6-21	10-11-12	3-4-5-9	13	1-3-5	3-5	10-11	12	8-12
R36	1-2	2-6	3-5	2-8-9-10	10	4-5-9	10-13-14	3	4	3-5-8	14	1-2-3-7-8
R40	1-2	2-6	2-3-7-9	2-20	1-3	5-9	10-12	3	3	5-6	2-6-7-11-14	2-8-12
R42	2	6	5-9-10	1-7-8-9-17	9	3-5-9	7-10-13	3-5	6	5-6-10	4-11-12-13	12
R43	2	6	6-7	11-12-15	9-10-13	3-5-8-9	10-13	1-3-5	3	2-5-6-10	8-12	8-12
R45	2	6	2-9	19	10	3-4-5-9	4-15	5-	3-5	5-9-10	14	12
R47	2	6	2-8-9	7-19	12	3-5-6-9	10-13	3	6	2-5-6-9	14	8-12
R48	2	8	2-8-9	2-7-9	6-10	4-5-9	9-10-13	3-7	1-3	5-6-7	14	8-10-11
R50	2	2-3-6	2-8-9	9-10-20	7-13	4-5-7-9	4-10-13	3	2	2-5-6	14	8
R52	2	4-6	2-5-9	1-9-11-13	2	5-8-9	11	3	2-4	2-5-8	5-9-11	8
R54	2	4-6	2-5-9	1-9-11-13	2-10-11	5-8-9	11	3	2-4-5	4	1-3-5-7-11	4-8
R57	2	4-6	2-5-9	1-9-11-13	2-10-11	5-7-8-9	11	3	2-4-5	2-5-8	1-5-7-11	4-8
R58	2	4-6	2-5-9	1-9-11-13	2-10-11	5-8-9	11	3	2-4-5	6	1-5-7-11	4-8
R59	2	4-6	2-5-9	1-9-11-13	2	5-7-8-9	11	3	2-4-5	1-2-5-6	1-5-7-11	4-8
R61	2	3-4-6	2-9	2-3-8-9-11-13	10-11	2-5-9	2	3	4-5	2-8-8-12	7-11	7-8
R67	2	6	2-4-9	16-19	10-11	5-9	4-13	3-5	4-5	5-6-10	14	8-12
R68	1-2	5-6	2-4-5-9	13-19	10-11	3-5-9	4-8-13	5-7	4-5	5-6-10	12	6-8-12
R71	2	6	2-3-9	5-13-19	12	5-9	8-14	3-5	4-5	5-10	12	12



ANNEXE 5 : Dendrogramme UPGMA élaboré à partir des distances de Jaccard calculées avec les données isoenzymatiques d'Ochman *et al.* (1984)

ANNEXE 6 : Tests statistiques D1 ET D2

test D1	génotypes	effectif observé	Fréquence attendue	Probabilité	Résultats	
<i>Escherichia coli</i> nos isoenzymes	R13	2	3,81 10 ⁻³	4,8 10 ⁻³	significatif	
	R36	2	5,81 10 ⁻⁴	1,17 10 ⁻⁴	significatif	
	R54	3	3,58 10 ⁻²	7,12 10 ⁻²	non significatif	
	isoenzymes d'Ochman	R13	3	7,3 10 ⁻⁹	1,1 10 ⁻⁹	significatif
		R52	2	3,7 10 ⁻⁴	4,9 10 ⁻⁵	significatif
	RAPD (12 amorces)	R52	2	2,49 10 ⁻⁴	2,17 10 ⁻⁵	significatif
RAPD combinaison A						
RAPD combinaison B						
<i>Trypanosoma cruzi</i> Isoenzymes	canIII	2	8,39 10 ⁻⁸	1,94 10 ⁻¹²	significatif	
	CBB	2	2,58 10 ⁻⁴	1,83 10 ⁻⁵	significatif	
	CUICA	3	5,4 10 ⁻⁴	3,17 10 ⁻⁷	significatif	
	M6241	3	4,31 10 ⁻⁶	1,62 10 ⁻¹³	significatif	
	NR	2	5,91 10 ⁻⁴	9,57 10 ⁻⁵	significatif	
	RAPD	79	4	2,95 10 ⁻⁴	8,05 10 ⁻¹¹	significatif
		CBB	2	6,85 10 ⁻⁴	1,28 10 ⁻⁴	significatif
		CUTIA	3	6,85 10 ⁻⁴	6,45 10 ⁻⁷	significatif

test D2	nb de génotypes répétés observés	nb de génotypes répétés attendus	Probabilité	Résultats	
<i>Escherichia coli</i> nos Isoenzymes	3	3	0,2123	non significatif	
		4	0,0273	significatif	
		5	0,03	significatif	
		6	0,002	significatif	
	isoenzymes d'Ochman	3	3	0,0009	significatif
	RAPD (12 amorces)	0			
RAPD combinaison A	3	1	1	non significatif	
RAPD combinaison B	0				
<i>Trypanosoma cruzi</i> isoenzymes	3	1	1	non significatif	
		2	0,0048	significatif	
	RAPD	4	1	1	non significatif
		2	0,0057	significatif	

ANNEXE 7 : Tests statistiques e et f

test e	nb de génotypes différents observés	nb de génotypes différents attendus	Probabilité	Résultats
<i>Escherichia coli</i> nos isoenzymes	23	20	0,0005	significatif
		21	0,0017	significatif
		22	0,0155	significatif
		23	0,0759	non significatif
	24	24	0,0001	significatif
	27			
	24	26	0,0031	significatif
isoenzymes d'Ochman RAPD (12 amorces) RAPD combinaison A	24	27	1	significatif
		27		
RAPD combinaison B	27			
<i>Trypanosoma cruzi</i> isoenzymes	17	23	0,0048	significatif
		24	1	significatif
	18	23	0,0057	significatif
		24	1	significatif

test f	moyenne du déséquilibre	déséquilibre attendu	Probabilité	Résultats
<i>Escherichia coli</i> nos Isoenzymes isoenzymes d'Ochman RAPD (12 amorces) RAPD combinaison A RAPD combinaison B	0,0007	0,001	0,05	non significatif
	0,0014	0,003	< 10 ⁻⁴	significatif
	0,0001	0,003	< 10 ⁻⁴	significatif
	0,0001	0,005	< 10 ⁻⁴	significatif
	0,0001	0,001	0,1555	non significatif
<i>Trypanosoma cruzi</i> isoenzymes RAPD	0,0019	0,021	< 10 ⁻⁴	significatif
	0,0019	0,017	< 10 ⁻⁴	significatif

ANNEXE 8 : tests statistiques g et g intragroupes

test g	coefficient de corrélation	Probabilité	Résultats
<i>Escherichia coli</i>			
nos Isoenzymes-RAPD(12 amorces)	0,44	< 10 ⁻⁴	significatif
isoenzymes d'Ochman-RAPD(12 amorces)	0,39	< 10 ⁻⁴	significatif
nos isoenzymes-RAPD combinaison A	0,46	< 10 ⁻⁴	significatif
nos Isoenzymes-RAPD combinaison B	0,28	1,4 10 ⁻³	significatif
<i>Trypanosoma cruzi</i>			
isoenzymes-RAPD	0,89	< 10 ⁻⁴	significatif

test g intragroupe	coefficient de corrélation	Probabilité	Résultats
<i>Escherichia coli</i>			
nos isoenzymes-RAPD(12 amorces)			
groupe 1	(-)0,08	0,82	non significatif
groupe 2	0,39	< 10 ⁻⁴	significatif
isoenzymes d'Ochman-RAPD(12 amorces)			
groupe 1	0,019	0,4427	non significatif
groupe 2	0,33	< 10 ⁻⁴	significatif
<i>Trypanosoma cruzi</i>			
isoenzymes-RAPD			
groupe 1	0,4	0,00157	significatif
groupe 2	0,74	< 10 ⁻⁴	significatif

ANNEXE 9

Test statistique f appliqué sans les génotypes répétés

test f	moyenne du déséquilibre	déséquilibre attendu	Probabilité	Résultats
<i>Escherichia coli</i>				
nos isoenzymes	0,0009	0,0009	0,4416	non significatif
isoenzymes d'Ochman	0,0015	0,0028	0,0004	significatif
RAPD combinaison A	0,001	0,002	0,0227	non significatif
<i>Trypanosoma cruzi</i>				
isoenzymes	0,002	0,018	< 10 ⁻⁴	significatif
RAPD	0,002	0,01	< 10 ⁻⁴	significatif

ANNEXE 10

Origine des 29 souches d'*E. coli* utilisées dans notre étude.

Données isoenzymatiques d'Ochman *et al.*(1984): Zymodèmes et groupes d'appartenance de ces souches.

N°	SOUCHE		ORIGINE	GROUPE	ENZYMES										
	ANCIENNE RÉF.	HOTE (sexe)			MDH	6PG	ADK	PEP2	GOT	IDH	GPI	ACON	MPI	G6PG	ADH
4	RM39A	HUMAIN (F)	IOWA	I	2	15	4	7	3	2	4	6	3	2	1
6	RM66C	HUMAIN (F)	IOWA	I	2	13	4	5	3	2	4	6	3	2	1
9	FN98	HUMAIN (F)	SUEDE	I	2	9	4	5	3	2	4	7	3	2	1
13	FN10	HUMAIN (F)	SUEDE	I	2	6	4	7	3	2	4	7	3	1	1
14	P62	HUMAIN (F)	SUEDE	I	2	6	4	7	3	2	4	7	3	1	1
15	FN13	HUMAIN (F)	SUEDE	I	2	6	4	5	3	5	4	7	8	2	0
16	RM191F(c)	LEOPARD	WASHINGTON	I	2	2	4	7	3	2	4	7	8	2	1
18	RM210F(d)	SINGE	WASHINGTON	I	2	6	4	7	3	2	4	7	3	1	1
22	RM215C	BOEUF	BALI	I	2	6	4	7	3	2	4	6	8	2	1
25	MSI	CHIEN	NEW-YORK	I	2	5	4	5	3	5	4	7	3	2	1
27	RM24J	GIRAFE	WASHINGTON	II	2	6	2	5	3	5	5	6	7	2	6
30	RM10A	BISON	ALBERTA	II	2	8	4	5	3	5	7	6	9	2	6
36	RM77B(b)	HUMAIN (F)	IOWA	II	6	2	4	5	3	5	7	7	7	2	6
40	P60	HUMAIN (F)	SUEDE	II	6	4	5	5	3	5	7	7	7	2	6
42	DAR1	HUMAIN (M)	MASSACHUSETTS	II	2	6	4	5	3	5	4	6	7	2	6
43	FN36	HUMAIN (F)	SUEDE	II	2	6	4	5	3	5	4	6	5	2	6
45	RM201C	COCHON	INDONESIE	II	2	6	4	5	3	5	7	6	8	2	6
47	RM211C	MOUTON	NELLE-GUINÉE	II	7	4	4	5	3	5	4	2	7	2	0
48	C90	HUMAIN (F)	SUEDE	II	2	6	4	5	4	5	6	6	7	2	1
50	P97	HUMAIN (F)	SUEDE	II	5	14	4	5	3	5	9	5	7	2	6
52	RM73A	OURANG-OUTAN	WASHINGTON	III	2	6	5	2	4	6	7	6	3	2	6
54	RM64A	HUMAIN	IOWA	III	2	6	5	2	3	5	7	6	3	2	6
57	RM71B	GORILLE	WASHINGTON	III	2	15	5	2	3	5	7	6	3	2	6
58	RM185S	LION	WASHINGTON	III	2	6	4	2	4	5	7	7	8	2	2
59	SIL8	HUMAIN (M)	MASSACHUSETTS	III	2	6	4	4	3	5	7	6	3	2	6
61	FN23	HUMAIN (F)	SUEDE	III	2	14	5	4	3	5	7	6	3	2	0
67	RM217T	CHEVRE	INDONESIE	III	2	10	4	5	3	5	7	6	3	2	6
68	RM224H(f)	GIRAFE	WASHINGTON	III	2	6	4	4	3	5	4	6	7	2	6
71	ABU84	HUMAIN (F)	SUEDE	III	2	5	4	7	3	5	2	6	3	2	0

ANNEXE 11

Dendrogrammes UPGMA élaborés à partir des données isoenzymatiques et RAPD de Tibayrenc *et al.* (1993) sur les 24 stocks de *T. cruzi*.

