



HAL
open science

Qualification de la vraisemblance des informations publiées dans la littérature scientifique - Etude d'un cas concret : 'Candidatus Phytoplasma pyri' et Pear decline

Myriam Dulor

► To cite this version:

Myriam Dulor. Qualification de la vraisemblance des informations publiées dans la littérature scientifique - Etude d'un cas concret : 'Candidatus Phytoplasma pyri' et Pear decline. Informatique et langage [cs.CL]. 2024. hal-04662925

HAL Id: hal-04662925

<https://irsn.hal.science/hal-04662925v1>

Submitted on 26 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License



Rapport de stage

Qualification de la vraisemblance des informations publiées dans la littérature scientifique

Etude d'un cas concret :

'*Candidatus Phytoplasma pyri*' et Pear decline

Auteure : Myriam DULOR

Encadrants de stage : Claire Nédellec, Nicolas Sauvion,
Robert Bossy, Delphine Battistelli

Master 2 Traitement automatique des langues - 2023-2024

Résumé

Toute publication scientifique relate les découvertes de pairs pour ajouter ensuite de nouvelles connaissances scientifiques dans le domaine considéré. Ce stage de trois mois (deux mois au moment de la soutenance) vise à déterminer comment qualifier la véracité des connaissances publiées dans la littérature scientifique. Ce travail s'inscrit dans le cadre plus large d'un projet de thèse sur la qualité et la nouveauté d'informations.

L'étude a porté spécifiquement sur la connaissance des interactions biologiques au sein des triplets (ou pathosystèmes) plantes-phytoplasmes-insectes vecteurs, notamment sur le pathosystème poirier-phytoplasme-vecteurs et les espèces d'insectes responsables de la transmission de la maladie du dépérissement du poirier en France et en Europe.

L'objectif est donc d'identifier et de proposer des indices bibliographiques, biologiques et linguistiques pour optimiser efficacement l'extraction automatique de l'information (EI), à partir d'un corpus de documents au format pdf sur un organisme modèle (ex. agent pathogène '*Candidatus Phytoplasma pyri*'). Les modalités de citations bibliographiques (ex. négation, spéculation, incertitude) seront aussi à analyser.

La méthodologie mise en place pour cela repose dans un premier temps sur la définition du sujet scientifique, intégrant les connaissances pertinentes à rechercher. Dans un second temps, diverses pistes de traitement automatique des langues (TAL) et d'EI ont été explorées, en particulier avec les solutions Alvis NLP (de l'équipe Bibliome) pour l'EI et Grobib (de Patrice Lopez, Science-miner) pour le pré-traitement de documents au format pdf. Pour identifier des indices bibliographiques, biologiques et linguistiques, deux cas d'étude ont été considérés. Ce rapport fait un bilan des résultats obtenus et présente quelques perspectives.

Remerciements

Je remercie en premier lieu Claire Nédellec et Nicolas Sauvion, mes encadrants de stage à l'INRAE, de m'avoir donné l'opportunité de faire ce stage au sein de leurs équipes. Je leur suis très reconnaissante d'avoir adapté l'offre de stage initiale à mon profil professionnel. J'ai eu le privilège d'avoir avec eux des échanges de grande qualité du point de vue scientifique et pédagogique, constructifs et très riches tout au long du stage.

Je remercie Robert Bossy pour les séances de formation à son outil Alvis NLP et de sa disponibilité pour régler des points techniques ponctuels tout au long du stage. Merci aussi à Patrice Lopez, auteur de l'outil Grobid, de sa disponibilité à répondre à des points techniques concernant son outil.

Je remercie l'ensemble de l'équipe des formateurs du Master TAL et de l'Université de Paris Nanterre, en particulier Delphine Battistelli pour sa pédagogie et son soutien dans le cadre de ces études et le stage.

Je suis reconnaissante à l'équipe IRSN (Flavien Lemoine, Jean-Paul Daubard, le Scope et la RH) de m'avoir accordé d'entamer un deuxième Master 2 pro plus axé sur les nouvelles technologies, après toutes ces années de travail au sein de l'IRSN.

Mes remerciements vont aussi à mon entourage, en particulier ma famille et mes proches, qui ne cessent de me soutenir et de m'encourager à persévérer. A Celui qui me fortifie en toutes choses.

Table des matières

Introduction générale.....	1
1) Contexte sociétal et scientifique du sujet.....	1
1.1) INRAE, institut de recherche et d'expertise.....	1
1.3) Phytoplasmes des arbres fruitiers & risque phytosanitaire.....	2
2) Problématique du stage et méthodologie.....	2
3) Focus sur l'équipe encadrante.....	3
Chapitre 1 Connaissances actuelles du contexte scientifique applicatif.....	5
1) Maladie du Pear decline & 'Candidatus Phytoplasma pyri'.....	5
2) Biologie de l'organisme modèle vecteur.....	9
3) Biologie des organismes "proches".....	13
Chapitre 2 Perspectives TAL.....	16
1) Constitution d'un corpus modèle.....	16
2) Text-mining avec l'outil AlvisNLP.....	17
1.1) Généralités.....	17
1.2) Présentation du script utilisé.....	19
1.3) Expérience avec mes propres dictionnaires.....	21
3) Problématique de conversion du format PDF au texte brut.....	22
3.1) TikaReader.....	22
3.2) Grobid.....	23
Chapitre 3 Vérification de l'information primaire : présentation de deux cas d'étude.....	26
1) Cas d'étude illustré avec la publication de Lemoine (1991).....	26
1.1) Introduction à ce cas d'étude : analyse conjointe d'une publication d'une auteure de référence, experte reconnue au niveau international.....	26
1.2) Accessibilité au document et à l'information contenue dans la publication de Lemoine (1991).....	27
1.3) Connaissance scientifique acquise avec cet article.....	29
1.4) Analyse des citations de la référence bibliographique (Lemoine, 1991).....	29
2) Cas d'étude illustré avec l'espèce " <i>Cacopsylla pyrisuga</i> ".....	33
2.1) Accessibilité aux documents cités et aux informations contenues dans ces publications.....	36
2.2) Connaissance scientifique acquise avec ces articles.....	37
2.3) Analyse des citations.....	38
Chapitre 4 Discussion & Conclusion.....	42
Synthèse concernant l'accès aux informations primaires citées.....	42
Synthèse concernant les indices bibliographiques, biologiques & linguistiques.....	45

Conclusion personnelle.....	49
Bibliographie.....	50
ANNEXE n°1 - Organigramme INRAE et présentation de l'unité de recherche MaIAGE....	53
ANNEXE n°2 - Vecteurs <i>Cacosylla pyri</i>, <i>C. pyricola</i>, <i>C. pyrisuga</i>.....	55
ANNEXE n°3 - Script Alvis NLP "main.plan" que j'ai fréquemment utilisé.....	59
ANNEXE n°4 - Dictionnaire du phytoplasme considéré et des divers nommages.....	62
ANNEXE n°5 - Dictionnaire des psylles considérés et des divers nommages.....	65
ANNEXE n°6 - Dictionnaire sur la maladie et ses divers nommages.....	73
ANNEXE n°7 - Dictionnaire sur la plante hôte et ses divers nommages.....	74
ANNEXE n°8 - Dictionnaire de verbes et marqueurs cités et intéressants dans le cadre de cette étude.....	78
ANNEXE n°9 - Script conçu sur la base des dictionnaires personnalisés.....	83
ANNEXE n°10 - Analyse détaillée concernant l'espèce <i>Cacosylla Pyrisuga</i>.....	86
ANNEXE n°11 - Tableau d'analyse de quelques indices linguistiques.....	100

Introduction générale

Ce rapport de stage de Master 2 Traitement automatique des langues (TAL) relate les riches expériences faites en deux mois pour aborder le sujet complexe de la qualification de la véracité des connaissances publiées dans la littérature scientifique, avec l'exemple des interactions biologiques au sein du pathosystème poirier-phytoplasme-vecteurs. Ce chapitre détaille le contexte et la problématique du sujet donné. Mon stage s'est déroulé à partir du 2 avril 2024, avec quatre jours de présence par semaine, au sein de l'organisme INRAE de Jouy-en-Josas (équipe MaLAGE) et en interaction à distance avec une équipe INRAE de Montpellier (équipe FORISK). Ces deux équipes sont présentées ci-après.

1) Contexte sociétal et scientifique du sujet

1.1) INRAE, institut de recherche et d'expertise

L'Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), est un acteur français majeur de la recherche et de l'innovation dans le domaine des sciences agronomiques et alimentaires, des sciences du végétal et de l'animal, et en écologie-environnement. L'institut rassemble une communauté de 12 000 personnes, avec 272 unités de recherche, de service et d'expérimentation implantées dans 18 centres sur toute la France (voir l'organigramme en [ANNEXE n° 1](#)). L'INRAE se positionne parmi les tout premiers organismes de recherche au monde spécialisé sur l'ensemble « agriculture-alimentation-environnement ». L'INRAE a pour ambition d'être un acteur clé des transitions nécessaires pour répondre aux grands enjeux mondiaux. Face à l'augmentation de la population et au défi de la sécurité alimentaire, au dérèglement climatique, à la raréfaction des ressources et au déclin de la biodiversité, l'Institut a un rôle majeur pour construire des solutions et accompagner la nécessaire accélération des transitions agricoles, alimentaires et environnementales. L'Institut a pour vocation de produire et diffuser des connaissances scientifiques sur ces sujets et de mobiliser ces connaissances au service de l'innovation, de l'expertise et de l'appui aux politiques publiques.

L'INRAE participe aux expertises scientifiques collectives indépendantes et pluralistes, en étant directement mandaté par les autorités françaises ou européennes. L'organisme peut être sollicité par exemple sur un des sujets comme la sécurité sanitaire dans les domaines de l'environnement et de l'alimentation et pour évaluer les risques sanitaires qu'ils peuvent comporter. L'INRAE peut être aussi sollicité pour ces expertises par l'ANSES (*agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail*). Ses avis sont rendus publics et ces expertises scientifiques contribuent à l'élaboration de directives à l'échelle de l'Union Européenne et d'arrêtés à l'échelle nationale. Dans ce cadre-là, l'Institut a participé à des expertises collectives sur les risques épidémiques, en particulier les risques associés à **des**

bactéries (phytoplasmes) d'arbres fruitiers transmises par des insectes piqueurs suceurs (Hémiptères) qui ont fait l'objet de mon stage.

1.3) Phytoplasmes des arbres fruitiers & risque phytosanitaire

La multiplication et la production d'espèces majeures pour l'arboriculture française, comme les espèces fruitières à noyaux (abricotiers) ou à pépins (pommiers et poiriers), et plusieurs de leurs porte-greffes, sont en situation délicate pour leur multiplication et leur production du fait de la présence d'un certain nombre d'agents pathogènes (virus, bactéries ou champignons).

En 2011, l'Anses a été saisie par la Direction générale de l'alimentation (DGAL) du Ministère en charge de l'agriculture, d'une demande d'avis relatif à la réalisation d'une analyse de risque phytosanitaire sur les phytoplasmes des arbres fruitiers. **Les phytoplasmes sont des bactéries qui causent des maladies d'arbres fruitiers** dont les impacts économiques sont très importants en Europe et dans le monde ([ANSES, 2021](#)). Trois analyses de risque phytosanitaire portant sur 3 phytoplasmes ont été réalisées. Ces phytoplasmes sont :

- '*Candidatus* Phytoplasma pyri', agent pathogène de la maladie du dépérissement du poirier (Pear decline ou PD)
- '*Candidatus* Phytoplasma mali', agent pathogène de la maladie de la prolifération du pommier (Apple Proliferation ou AP)
- '*Candidatus* Phytoplasma prunorum', agent pathogène de la maladie de l'European Stone Fruit Yellows (ESFY)

Ces bactéries pathogènes s'attaquent à différents types de plantes de la famille des Rosacées (poiriers, pommiers, *Prunus*, pommiers) qui ont de forts impacts économiques en arboriculture. Les bactéries sont transmises d'une plante à l'autre par des insectes piqueurs suceurs (ordre des Hémiptères) apparentés aux pucerons, des psylles du genre *Cacopsylla*.

La partie scientifique de mon stage de Master 2 TAL a porté spécifiquement sur le triplet psylle vecteur-poirier-'*Candidatus* Phytoplasma pyri'.

2) Problématique du stage et méthodologie

La connaissance des interactions biologiques au sein des triplets (ou pathosystèmes) insectes vecteurs-plantes-phytoplasmes est incomplète et mal établie, notamment en raison du grand nombre de publications scientifiques dispersées depuis le XIX siècle. Pour le pathosystème psylles vecteurs-poirier-phytoplasme, les espèces d'insectes responsables de la transmission restent à clarifier.

L'objet de mon stage a porté sur l'Extraction d'Information (EI) automatique dans un corpus de publications scientifiques, sur la maladie du Pear decline.

L'objectif général était de **qualifier par des approches automatiques la vraisemblance et l'originalité des informations publiées dans la littérature scientifique.**

A partir d'un corpus de documents sur un organisme modèle (ex. le psylle vecteur *Cacopsylla pyri*), les objectifs initiaux étaient :

- d'identifier des indices bibliographiques, linguistiques et biologiques pertinents,
- de réaliser l'état de l'art en traitement automatique des langues (TAL) sur l'analyse de modalités de citations bibliographiques (ex. Négation, Spéculation, Incertitude),
- d'utiliser et concevoir des workflows d'extraction d'information au moyen de la plateforme Alvis NLP de l'équipe Bibliome,

de réaliser des comptes-rendus réguliers appuyés sur des exemples et animer les échanges avec l'équipe encadrante.

3) Focus sur l'équipe encadrante

Dans le cadre de ce stage, j'étais membre de l'équipe Bibliome.

L'équipe Bibliome est rattachée à l'Unité de recherche [Mathématiques et Informatique Appliquées du Génome à l'Environnement \(MaIAGE\)](#) (voir [ANNEXE n° 1](#) pour plus de détails). L'**équipe Bibliome** est spécialisée dans l'acquisition et formalisation de connaissances à partir de textes. Elle développe des méthodes de traitement du langage naturel et d'apprentissage automatique pour extraire des informations de textes dans le domaine de la biologie. L'équipe travaille sur des tâches spécifiques d'extraction d'information (EI) telles que la reconnaissance d'entités, la normalisation d'entités (*entity linking*) et l'extraction de relations. Elle se concentre sur les méthodes qui combinent l'information linguistique, l'apprentissage automatique et la connaissance du domaine (ontologies et taxonomies) et qui sont capables d'apprendre à partir d'un petit nombre d'exemples. Ces méthodes sont appliquées à un large éventail d'applications en Sciences de la Vie - de la diversité microbienne à la biologie végétale et à la surveillance épidémiologique. Une part importante de l'activité consiste également à promouvoir le développement et l'évaluation de systèmes d'EI en organisant des *challenges*.

L'équipe Bibliome est constituée de 4 membres permanents, d'une post-doctorante, d'une étudiante en thèse et d'une stagiaire (moi-même) :

Claire Nédellec : directrice de recherche et responsable de l'équipe

Robert Bossy : ingénieur de recherche et responsable de la "suite Alvis"

Louise Deléger : chargée de recherche

Arnaud Ferré : chargé de recherche

Marine Courtine : post-doc

Mariya Borovikova : étudiante en thèse

Le sujet étant à la fois scientifique et sur les compétences de TAL, j'ai eu comme encadrants de stage Claire NÉDELLEC, Robert BOSSY et Nicolas SAUVION (N. Sauvion, épidémiologiste, entomologiste rattaché à l'unité de recherche PHIM-INRAE basée à Montpellier). Ces trois chercheurs-experts, chacun dans leur domaine, ont l'habitude de collaborer ensemble afin de faire avancer la science.

L'unité mixte de recherche (UMR) PHIM (*Plant Health Institute of Montpellier*), à laquelle Nicolas Sauvion est rattaché, est composée de 140 permanents (agents INRAE, CIRAD, IRD, Université Montpellier 2, SupAgro Montpellier). Cette unité mobilise des compétences en bactériologie, entomologie, mycologie, virologie, épidémiologie, génétique des populations, phytopathologie, biologie moléculaire et bio-informatique. Durant la période de mon stage, j'ai collaboré avec deux stagiaires M1 et M2 biologistes qui effectuaient leur stage avec N. Sauvion au même moment que moi sur le même objet d'étude (psylles vecteurs) mais pour répondre à des questions différentes.

Chapitre 1 Connaissances actuelles du contexte scientifique applicatif

Lors des premières semaines de mon stage, j'ai découvert et assimilé le contexte scientifique de mon stage, en échangeant régulièrement avec N. Sauvion et ses stagiaires. De plus, dès le début, N. Sauvion a mis à ma disposition sa bibliothèque de références scientifiques sur la famille des insectes considérés, disponible sous Zotero et, un corpus composé à l'origine d'environ vingt-cinq publications scientifiques "phares" sur le sujet des interactions biologiques au sein du pathosystème vecteurs-poirier-phytoplasme sous format pdf.

1) Maladie du Pear decline & 'Candidatus Phytoplasma pyri'

La maladie du **Pear decline** affecte les poiriers et l'agent causal est le phytoplasme '**Candidatus Phytoplasma pyri**'. Cet organisme, du fait de sa nuisibilité, est pris en compte à l'échelle de l'Union européenne, notamment dans le cadre de la directive 2000/29/CE ([Conseil de l'Union européenne, 2000](#)). En France, il était classé "danger sanitaire de catégorie 2" au titre de l'arrêté interministériel du 15 décembre 2014 ([République Française, 2014](#)).

d) Virus et organismes analogues

1. Mycoplasme de la prolifération du pommier (Apple proliferation mycoplasma)
2. Mycoplasme de l'enroulement chlorotique de l'abricotier (Apricot chlorotic leafroll mycoplasma)
3. Mycoplasme du dépérissement du poirier (Pear decline mycoplasma)

Fig n°1 : Extrait de la Directive 2000, avec mention en point 3 du virus "Mycoplasme du dépérissement du poirier Pear decline mycoplasma"

Phytoplasme	Candidatus Phytoplasma pyri
-------------	-----------------------------

Fig n°2 : Quatre ans plus tard, extrait de l'arrêté interministériel 2014, avec mention du phytoplasme (bactérie) 'Candidatus Phytoplasma pyri' (notion eq. à Mycoplasme du dépérissement du poirier cité dans la figure n°1)

Depuis le 14 décembre 2019, ce phytoplasme est régulé à l'échelle de l'UE dans le cadre du règlements 'Santé des végétaux' [Règlement \(UE\) 2016/2031](#) et [Règlement d'exécution \(UE\) 2019/829](#). De "Organisme de quarantaine" (OQ), il est devenu "Organisme réglementé non de quarantaine" (ORNQ). Toutefois, le système d'autorisation de quarantaine est toujours en vigueur pour cet organisme en France, car il est susceptible d'être réglementé de façon complémentaire au niveau national.

La base de données européenne de référence *EPPO Global Database*, maintenue par le Secrétariat [European and Mediterranean Plant Protection Organization \(EPPO\)](http://www.eppo.org) et servant de base pour l'aide à la décision publique, donne les informations ci-après sur le phytoplasme '*Candidatus Phytoplasma pyri*'. On y découvre en particulier des synonymes, par exemple "16SrX-C" pour le phytoplasme considéré, ainsi que des informations importantes sur ses plantes hôtes, le nom de la maladie causée, ses vecteurs, etc.

'Candidatus Phytoplasma pyri' (PHYPPY) 📄 🌐 🐦

Code created in: 2005-03-22

MENU

- Overview --
- Distribution
- Host plants
- Host commodities
- Vectors
- Categorization
- Reporting
- Photos
- Documents
- Datasheet
- EPPO links

Overview

Basic information

- **EPPO Code:** PHYPPY
- **Preferred name:** '*Candidatus Phytoplasma pyri*'
- **Authority:** Seemüller & Schneider

Notes

16Sr X-C

Other scientific names

Name	Authority
Pear decline	
phytoplasma	
Phytoplasma pyri	Seemüller & Schneider

Taxonomy

- Kingdom: Bacteria (1BACTK)
- Phylum: Tenericutes (1TENEP)
- Class: Mollicutes (1MOLLC)
- Order: Acholeplasmatales (1ACHOO)
- Family: Acholeplasmataceae (1ACHOF)
- Genus: Phytoplasma (1PHYPG)
- Species: '*Candidatus Phytoplasma pyri*' (PHYPPY)

Common names

Name	Language
<input type="text" value="Search..."/>	- select -
decline of pear	English
leaf curl of pear	English
moria disease of pear	English
Parry's disease of pear	English
PD	English
dépérissement du poirier	French
flétrissement du poirier	French
Birnbaumsterben	German
Birnenverfall	German
fitoplasma do declínio da pereira	Portuguese (PT)
Фитоплазма истощения груши	Russian
decaimiento del peral	Spanish

Fig n°3 : Informations sur le phytoplasme '*Candidatus Phytoplasma pyri*' extraites d'*EPPO Global Database* (<https://gd.eppo.int/taxon/PHYPPY>), base de données de référence dans le domaine d'expertise considéré.

Avec l'information mentionnée "16SrX-C" dans cette fiche EPPO, je découvre grâce à l'aide de N. Sauvion que le phytoplasme qui m'intéresse est rattaché aux espèces " 'Candidatus Phytoplasma' " et au groupe X (10) "16SrX" ou "apple proliferation group".

Une autre source d'information intéressante (car elle fait référence en biologie) est la base de données [NCBI Taxonomy](#) mise en ligne par le National Library of Medicine aux Etats-Unis. Selon cette base de données, la lignée du phytoplasme 'Candidatus Phytoplasma pyri' est [Bacteria](#); [Mycoplasmata](#); [Mollicutes](#); [Acholeplasmatales](#); [Acholeplasmataceae](#); [Candidatus Phytoplasma](#); [16SrX \(Apple proliferation group\)](#).

NCBI Taxonomy Browser

Search for: [] AS complete name [] lock [] Go [] Clear

Display: 3 levels using filter: none

Candidatus Phytoplasma pyri

Taxonomy ID: 47566 (for references in articles please use NCBI:txid47566)

current name

"Candidatus Phytoplasma pyri" Seemuller and Schneider 2004, candidatus name¹

equivalent: **Phytoplasma pyri**

NCBI BLAST name: mycoplasmas

Rank: **species**

Genetic code: [Translation table 11 \(Bacterial, Archaeal and Plant Plastid\)](#)

Other names:

heterotypic synonym

Pear decline mycoplasma-like organism

heterotypic synonym

Pear decline phytoplasma

Lineage(abbreviated)

[Bacteria](#); [Mycoplasmata](#); [Mollicutes](#); [Acholeplasmatales](#); [Acholeplasmataceae](#); [Candidatus Phytoplasma](#); [16SrX \(Apple proliferation group\)](#)

Notes:

1) A provisional name for well characterised but as-yet uncultured organisms.

Comments and References:

[Murray R.G. & Schleifer, K.H. \(1994\)](#)
Murray, R.G., and Schleifer, K.H. "Taxonomic notes: a proposal for recording the properties of putative taxa of prokaryotes." *Int. J. Syst. Bacteriol.* (1994) 44(1):174-176.

[Murray R.G. & Stackebrandt, E. \(1995\)](#)
Murray, R.G., and Stackebrandt, E. "Taxonomic note: implementation of the provisional status Candidatus for incompletely described prokaryotes." *Int. J. Syst. Bacteriol.* (1995) 45(1):186-187.

[Seemuller E. & Schneider B. \(2004\)](#)
Seemuller, E., and Schneider, B. "Candidatus Phytoplasma mali" sp. nov., 'Candidatus Phytoplasma pyri' sp. nov. and 'Candidatus Phytoplasma prunorum' sp. nov., the causal agents of apple proliferation yellows, respectively." *Int. J. Syst. Evol. Microbiol.* (2004) 54:1217-1226.

Fig n°4 : Informations sur le phytoplasme 'Candidatus Phytoplasma pyri' extraites du NCBI taxonomy (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=47566>), autre base de données de référence dans le domaine d'expertise considéré.

La maladie du Pear decline affecte les poiriers, *Pyrus spp.* D'après [NCBI Taxonomy](#), la taxonomie de *Pyrus* est [cellular organisms](#); [Eukaryota](#); [Viridiplantae](#); [Streptophyta](#); [Streptophytina](#); [Embryophyta](#); [Tracheophyta](#); [Euphyllophyta](#); [Spermatophyta](#); [Magnoliopsida](#); [Mesangiospermae](#); [eudicotyledons](#); [Gunneridae](#); [Pentapetalae](#); [rosids](#); [fabids](#); [Rosales](#); [Rosaceae](#); [Amygdaloideae](#); [Maleae](#)

Lineage (full): [cellular organisms](#); [Eukaryota](#); [Viridiplantae](#); [Streptophyta](#); [Strept](#); [Pentapetalae](#); [rosids](#); [fabids](#); [Rosales](#); [Rosaceae](#); [Amygdaloideae](#); [Maleae](#)

- o **Pyrus** (pears) *Click on organism name to get more information.*
 - [Pyrus amygdaliformis](#)
 - [Pyrus argyrophylla](#)
 - [Pyrus betulifolia](#)
 - [Pyrus boissieriana](#)
 - [Pyrus bourgaeana](#)
 - o [Pyrus calleryana](#)
 - [Pyrus calleryana var. lanceata](#)
 - [Pyrus chosrovica](#)
 - o [Pyrus communis](#) (pear)
 - [Pyrus communis subsp. caucasica](#)
 - [Pyrus communis subsp. communis](#)
 - [Pyrus communis x Pyrus pyrifolia](#)
 - [Pyrus communis x Pyrus ussuriensis](#)

Fig n°5 : Informations sur la plante hôte *Pyrus spp.* extraites du NCBI taxonomy (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=3766&lvl=3>)

La maladie du Pear decline est caractérisée par des rougissements et dépérissements des poiriers affectés et son évolution s'étale sur quelques années.



Fig n°6 : Photos illustrant l'évolution de la maladie provenant du site web EPPO déjà mentionné précédemment

Des arbres porteurs du phytoplasme et n'exprimant aucun symptôme peuvent être présents en fortes proportions dans les vergers. Depuis quelques années, une recrudescence des symptômes et des cas de contaminations en pépinière et en vergers est observée.

Cette maladie est épidémique¹ sur le territoire français, avec des cas de contaminations recensés depuis plus d'un siècle. Déduire le processus de dispersion des agents pathogènes des plantes à transmission vectorielle, en particulier au début de l'émergence de la maladie, est un grand défi, nécessaire pour limiter le recours à des traitements pesticides.

¹ Qui a les caractères d'une épidémie (p. oppos. à *endémique* et à *sporadique*)

2) Biologie de l'organisme modèle vecteur

Comprendre le fonctionnement d'une épidémie nécessite d'acquérir de nombreuses connaissances sur l'identité exacte des organismes impliqués dans le pathosystème : quelle(s) plante(s) hôte(s) ? quelle(s) espèce(s) d'insectes vectrice(s) de quel agent pathogène ? quelle est la répartition géographique de la maladie ? etc. Recueillir ces informations n'est pas trivial et demande souvent plusieurs années d'observations et d'analyses..

Les insectes vecteurs de phytoplasmes des arbres fruitiers (Rosacées) appartiennent au genre *Cacopsylla*, qui comprend à ce jour 130 espèces essentiellement sur arbres et arbrisseaux ([Sauvion, 2022](#)).

Dans le cas présent, selon l'EPPO, les vecteurs de l'agent pathogène '*Candidatus Phytoplasma pyri*' en France et en Europe sont tous des insectes piqueurs de l'ordre Hemiptera, de la super-famille des Psylloidea (ou psylles) et associés au genre *Cacopsylla*, en l'occurrence : ***Cacopsylla pyri*** (Linnaeus, 1758), ***Cacopsylla pyricola*** (Forster, 1848), ***Cacopsylla pyrisuga*** (Foerster, 1848) (voir plus des détails sur ces espèces dans l'[ANNEXE n° 2](#)).

Organism	Type
<i>Cacopsylla pyri</i> (PSYLPi)	Known vector
<i>Cacopsylla pyricola</i> (PSYLPc)	Known vector
<i>Cacopsylla pyrisuga</i> (PSYLPy)	Known vector

Fig n°7 : Informations sur les insectes-vecteurs de '*Candidatus Phytoplasma pyri*' en France & Europe extraites d'EPPO Global Database (<https://gd.eppo.int/taxon/PHYPPY/vectors>)

Après échanges avec N. Sauvion, j'ai pu clarifier quelques notions et termes biologiques importants :

- Un insecte vecteur ne transmet pas la maladie: un insecte vecteur transmet un agent pathogène qui induit une maladie.
- Un psylle (insecte) est considéré comme un vecteur de bactérie pathogène (ici phytoplasme) si, après l'avoir acquis en se nourrissant du phloème d'une plante hôte infectée, il est ensuite capable de l'inoculer à une plante hôte.
- Un psylle peut acquérir la bactérie pathogène en s'alimentant de sève. Il devient ainsi porteur. Pour autant cela ne démontre pas qu'il est vecteur. Seule la preuve expérimentale de sa capacité à inoculer ensuite l'agent pathogène pourra confirmer son statut de vecteur.

- Une maladie peut se propager entre parcelles cultivées mais aussi peut provenir de l'environnement sauvage (ex. poiriers sauvages). Les connaissances actuelles ne permettent pas de dire clairement si des espèces de psylles connues sur poiriers sauvages peuvent être des sources de contamination des poiriers cultivés.
- Suite à nos discussions, nous avons défini trois classes de vecteurs : avéré, potentiel (dans l'attente de devenir avéré ou rejeté), rejeté (il n'est pas exclu d'imaginer qu'un vecteur puisse passer d'un statut avéré à rejeté).

Le modèle de connaissance du pathosystème de référence dans le cadre de mon stage est le suivant :

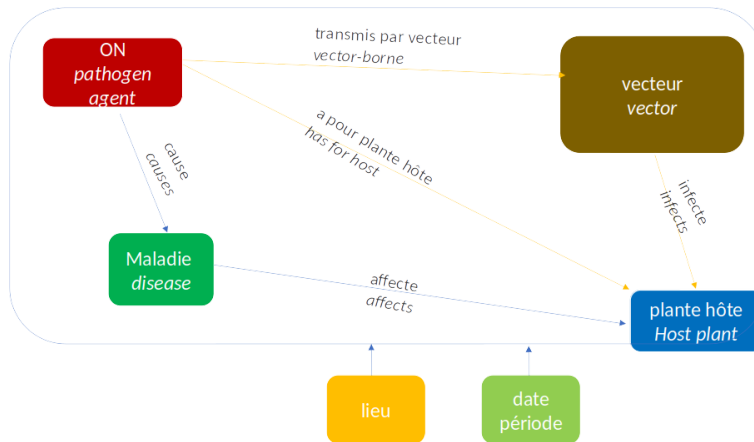


Fig n°8 : Graphe de connaissance de base d'un pathosystème impliquant un insecte vecteur, relations n-aires. L'agent pathogène (phytoplasme) est transmis par un vecteur (insecte). Ce vecteur infecte la plante-hôte en inoculant l'agent pathogène dans les vaisseaux conducteurs de sève élaborée (= phloème) de la plante-hôte. L'agent pathogène est ici spécifique à une plante hôte. L'agent pathogène cause une maladie chez la plante hôte. Cette maladie se traduit par divers symptômes chez la plante hôte. Ce pathosystème est généralement spécifique à un lieu géographique et une saison dans l'année.

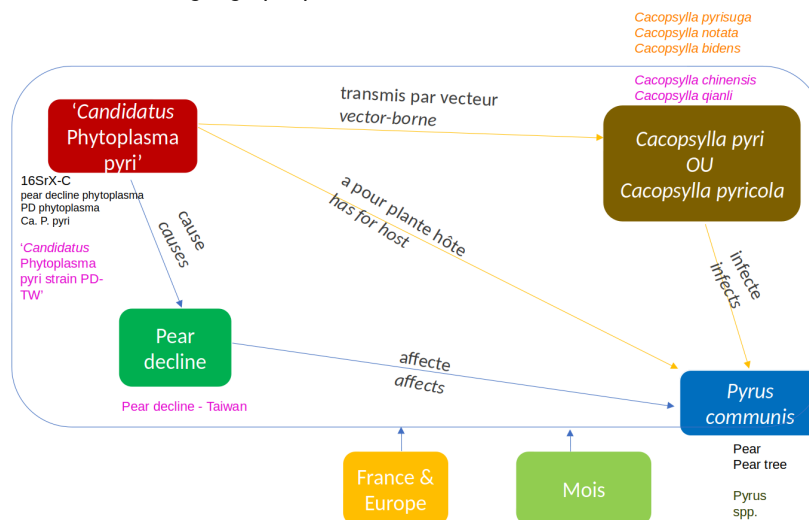


Fig n°9 : Graphe de connaissance de référence de relations n-aires pour le pathosystème psylles-poirier-phytoplasme en Europe et en France en particulier.

Sur la figure n°9, le phytoplasme ‘*Candidatus Phytoplasma pyri*’ (avec une liste non exhaustive en noir de nommage similaire comme 16SrX-C, etc) apparaît comme transmis en France ou en Europe par les insectes-vecteurs *Cacopsylla pyri* ou *C. pyricola*. Ce phytoplasme a pour plante hôte les arbres fruits de type *Pyrus spp.* (famille du Poirier), principalement l’espèce de poirier *Pyrus communis*. Il cause la maladie du dépérissement du poirier, Pear decline. Ces faits sont certains (de nombreuses publications l’attestent comme par exemple la publication [\(Jarausch and al. 2019\)](#)) et validés par N. Sauvion.

Je me suis intéressée, nous le verrons plus loin en détail, au cas plus ambiguë de *Cacopsylla pyrisuga* (en orange), un vecteur considéré comme “avéré” par l’EPPO.

Deux autres espèces auraient mérité notre attention, *Cacopsylla notata* (Flor, 1861) et *Cacopsylla bidens* (Sulc, 1907) (notées en orange également). Mais faute de temps, nous n’avons pas pu étudier ces cas. Qui plus est, il existe très peu de connaissances sur ces deux espèces ; par conséquent, il subsiste aussi des doutes concernant leurs rôles de vecteurs en France et en Europe.

La couleur rose représente le pathosystème connu existant à ce jour en Asie : le phytoplasme ‘*Candidatus Phytoplasma pyri* strain PD-TW’ est transmis par les insectes-vecteurs *Cacopsylla chinensis* (Yang & Li, 1981) ou *Cacopsylla qianli* (Yang & Li, 1984). Même si aujourd’hui ce pathosystème est spécifique à l’Asie et n’existe pas encore en France et Europe, selon les experts français et européens il est important de rester en veille en prévision d’une éventuelle apparition sur nos territoires.

Les variétés d’espèces de *Pyrus* citées dans les publications, ainsi que l’association à des lieux géographiques, seraient intéressantes à prendre en compte dans un travail futur. Là aussi, nous ne l’avons pas fait faute de temps, mais ce type d’informations (association de certaines espèces de psylles à certaines espèces de *Pyrus*, dans certaines zones géographiques précises) pourraient être des critères utiles de validation de la connaissance. A l’heure actuelle, par exemple, les connaissances sur la répartition géographique des espèces de psylles du poirier est très imprécise, comme le démontre une figure d’un article de *review* très récent :

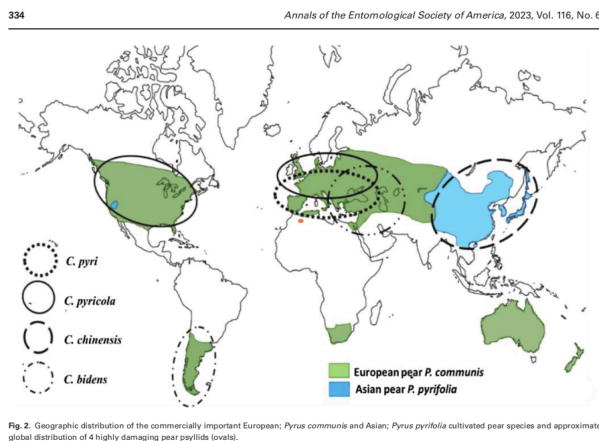


Fig n°10 : Distribution géographique des insectes vecteurs, extrait de la review de [Civolani and al. \(2023\)](#)

Au fil des lectures scientifiques (en particulier la *review* de [Civolani and al. \(2023\)](#) et des échanges avec N. Sauvion, j'ai pu aussi relever les caractéristiques suivantes du pathosystème considéré : les espèces d'insectes considérés, *C. pyri*, *C. pyricola* et *C. pyrisuga* ont chacune un cycle de vie saisonnier spécifique. Pendant l'hiver et la période de dormance (phase de vie ralentie), les psylles peuvent se réfugier sur des espèces de plantes différentes des poiriers, appelées plantes *refuges* (*shelter plants* en anglais). Il peut aussi y avoir des caractères morphologiques différents d'une saison à l'autre (été vs. hiver) ainsi que d'un genre à un autre (mâle vs. femelle). Je ne m'attarderai pas ici sur ces informations, mais je pense qu'il faudra en tenir compte dans le futur.

plante hôte # plante refuge (l'insecte s'y réfugie lorsque les conditions ne sont pas favorables pour lui)

cas envisageable : même insecte mais dimorphisme possible suivant la saison
(forme d'hiver et d'été)

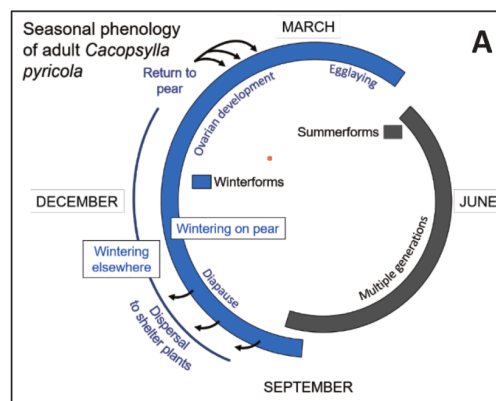


Fig n°11 : Cycle de vie de *C. pyricola*, extrait de la *review* de [Civolani and al. \(2023\)](#)

Derrière l'image assez simple du "triplet" se cachent en fait souvent une réalité beaucoup plus complexe : alternance de plantes hôtes cultivées et sauvages, compartiments cachés, mélange de l'agent pathogène à différentes échelles temporelles et spatiales en fonction des schémas de dispersion (souvent inconnus) des insectes vecteurs, disponibilité d'outils d'identification (par exemple, des marqueurs moléculaires) et de méthodes d'analyse appropriées, stratégies d'échantillonnage concevables à la lumière des connaissances préalables du pathosystème, etc. Maintenir à jour des connaissances demandent ainsi souvent une forte expertise et des efforts importants en terme de recherche d'informations et d'assimilations de faits nouveaux. Utiliser le TAL devrait permettre d'aider à cette mise à jour de toutes ces connaissances. Cependant, mon stage ne visait pas à aborder l'ensemble de ces points et je me suis restreint à une question plus spécifique comme cas d'étude : les trois espèces de psylles vectrices associées au Pear decline (*C. pyri*, *C. pyricola* et *C. pyrisuga*).

Dans le corpus de publications scientifiques analysé au préalable, j'ai remarqué un faible pourcentage de publications entièrement dédiées aux espèces de psylles vectrices qui m'intéressaient. Souvent, ces publications entièrement dédiées à ces espèces les mentionnent dès le titre ou dans le résumé de la publication. Ces publications relatent aussi généralement des expériences en laboratoire, ou in situ dans des vergers. Dans la majorité des publications considérées qui sont issues du corpus fourni, un état de l'art de connaissance

scientifique mondiale sur la famille ou le genre de l'espèce de psylle et les diverses maladies d'arbres fruitiers (*Rosacées*) est présenté sous forme de *review*. De ce fait, généralement, les données textuelles sur les espèces très spécifiques qui m'intéressent sont noyées dans la publication considérée. Si l'on raisonne en proportion de mots, cela équivaut presque à chercher une aiguille dans une botte de foin.

De plus, il apparaît clairement à l'issue des échanges avec N. Sauvion qu'il reste encore beaucoup d'inconnues sur la biologie et la classification (au sens taxonomique du terme) des ces organismes qui m'ont servi de modèle. De nouvelles données de connaissance apparaîtront dans le futur avec de nouvelles découvertes, validées par des données expérimentales.

3) Biologie des organismes "proches"

Au fil des échanges scientifiques, j'ai noté en marge les informations présentées ci-après concernant quelques autres pathosystèmes proches du mien, souvent citées dans les publications relatant les états de l'art. Les graphes de connaissance ci-après représentent une connaissance connexe, hors sujet par rapport au mien. Toutefois, il m'est apparu important d'avoir cette connaissance connexe afin de mieux comprendre les publications considérées et éventuellement de mieux détecter les anomalies ou les confusions de citations de connaissance. De plus, ces informations concernant des insectes "très proches" nous aident à mieux appréhender toute la complexité du sujet.

Le phytoplasme qui m'intéresse, '*Candidatus Phytoplasma pyri*', peut aussi être à l'origine d'une autre maladie, le *Peach yellow leaf roll*, chez un autre arbre fruitier, le pêcher (*Prunus persica* (L.) Batsch), ou pêcher commun. L'insecte vecteur est aussi *C. pyricola*, déjà citée. Ce cas existe aux Etats-Unis selon [Jarausch and al. \(2019\)](#).

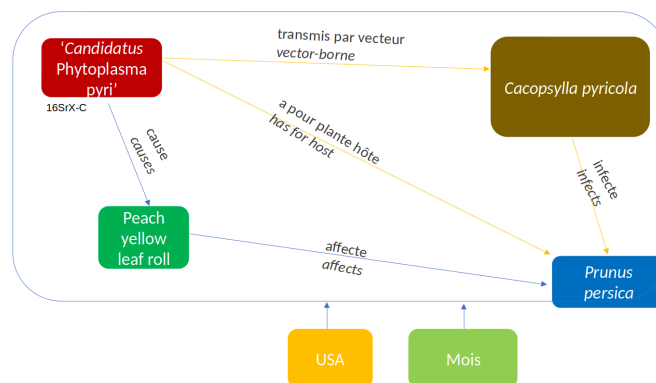


Fig n°12 : Graphe de connaissance concernant la maladie Peach yellow leaf roll, traduite en relations binaires. L'agent pathogène '*Candidatus Phytoplasma pyri*' est transmis par l'insecte-vecteur *Cacopsylla pyricola*. Ce vecteur infecte la plante-hôte en inoculant dans le phloème de la plante-hôte. Cet agent pathogène est spécifique à une plante hôte *Prunus persica* (L.) Batsch. L'agent pathogène cause la maladie du Peach yellow leaf roll chez la plante hôte. Cette maladie se traduit par divers symptômes chez la plante hôte. Ce pathosystème est généralement spécifique aux Etats-Unis. Dans le cadre de ce stage, je n'ai pas prêté attention à la saison.

Un autre phytoplasme ‘*Candidatus Phytoplasma mali*’, proche génétiquement de celui qui m’intéresse ‘*Candidatus Phytoplasma pyri*’, peut causer la maladie *Apple proliferation* au sein des pommiers. Ce phytoplasme est transmis par deux autres espèces d’insectes, *Cacopsylla picta* (Foerster, 1848) et *Cacopsylla melanoneura* (Foerster, 1848) selon [Jarausch and al. \(2019\)](#). *Cacopsylla picta* véhicule le phytoplasme de verger à verger selon N. Sauvion.

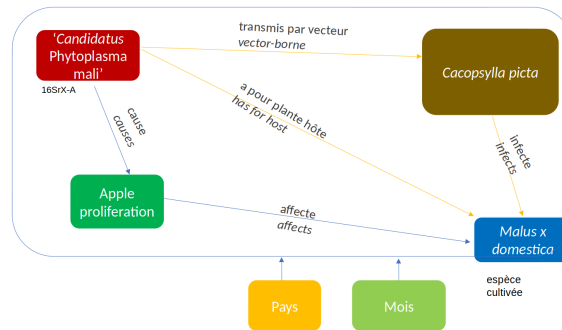


Fig n°13 : Graphe de connaissance illustrant les relations binaires au sein du pathosystème impliquant poiriers cultivés, l’agent pathogène et le psylle vecteur *Cacopsylla picta*.

Cacopsylla melanoneura est une espèce très présente sur aubépine. Cette plante peut être infectée par le phytoplasme trouvé sur le pommier selon [Jarausch and al. \(2019\)](#). Il est ainsi très vraisemblable que *C. melanoneura* soit à l’origine de transmission du milieu sauvage aux vergers selon N. Sauvion.

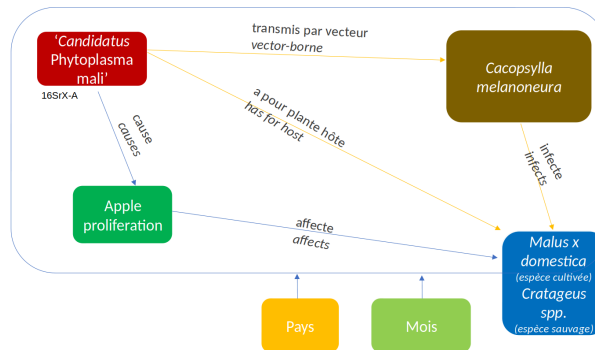


Fig n°14 : Graphe de connaissance illustrant les relations binaires au sein du pathosystème impliquant les poiriers cultivés, l’agent pathogène et le psylle vecteur *Cacopsylla melanoneura*.

Le phytoplasme ‘*Candidatus Phytoplasma prunorum*’ est à l’origine de l’*European stone fruit yellows* (ESFY), une maladie qui affecte des *Prunus*, en particulier les pruniers japonais et les abricotiers. Dans ce cas, le phytoplasme est transmis par un complexe de deux espèces de psylles, provisoirement appelées *Cacopsylla pruni* (Scopoli, 1763) espèce A et espèce B selon [Jarausch and al. \(2019\)](#). La maladie et les vecteurs sont endémiques en Europe dans les *Prunus* sauvages (type prunelliers), des plantes qui jouent ainsi un rôle essentiel dans la dispersion de la maladie au sein des vergers selon la publication ([Mariejeanne et al., 2020](#))

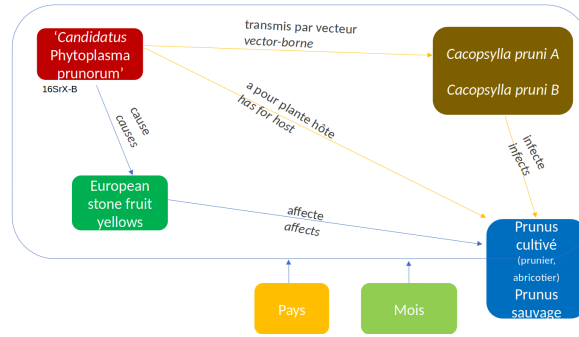


Fig n°15 : Graphe de connaissance illustrant les relations binaires au sein du pathosystème associé à la maladie sur Prunus, l'European stone fruit yellows.

Chapitre 2 Perspectives TAL

L'objet de mon stage porte sur l'Extraction d'Information (EI) automatique dans un corpus de publications scientifiques concernant le sujet présenté au chapitre précédent. L'objectif général est de qualifier par des approches automatiques la vraisemblance et l'originalité des informations publiées dans la littérature scientifique.

1) Constitution d'un corpus modèle

Pour ce projet, les données textuelles sont extraites de publications scientifiques disponibles sous format pdf.

Si l'on souhaite avoir une vue d'ensemble générale de la représentativité du nombre de publications scientifiques mondiales sur le sujet donné, on peut bien-sûr interroger [Google scholar](#) sur le sujet mais aussi les deux grandes bases de données bibliographiques multidisciplinaires internationales, à partir desquelles la science est évaluée par les Conseils de l'évaluation de la recherche, qui sont le [Web of Science](#) (WOS) de l'éditeur Clarivate analytics et [Scopus](#) de l'éditeur Elsevier et pour lesquelles l'INRAE a un abonnement.

J'ai fait un test avec la requête "**Pear decline**" OR ((**cacopsylla** OR **psyll***) AND **pear**) début mai et j'ai obtenu 546 références bibliographiques avec Scopus et 793 avec Web of Science pour une période allant de 1892 à aujourd'hui. En fusionnant les données et supprimant les doublons, j'obtiens 914 résultats avec un plus grand nombre de publications depuis les années 90.

Période	Proportion de références bibliographiques (calculée sur la base de données Scopus)
2020-2029	19,2 %
2010-2019	26,2 %
2000-2009	26,4 %
1990-1999	17,3 %
1960-1989	10,2 %
1761-1959	0,5 %

Fig n°16 : Répartition du nombre de références par an et exprimé en % (calculée sur la base des données Scopus)

En analysant ces références et en les comparant avec la liste des références compilées par Nicolas Sauvion, je note que ces données issues de Scopus et du WOS ne sont pas exhaustives sur le sujet.

Dans le cadre de ce projet, Nicolas Sauvion a lui-même identifié et isolé dans un dossier les publications scientifiques les plus pertinentes sur le sujet donné, comme déjà mentionné au

chapitre 1. Sa sélection s'est effectuée principalement sur la base de sa connaissance experte du sujet et de sa veille depuis des décennies. Sa bibliographie est composée à ce jour d'environ 150 références bibliographiques très pertinentes. En parallèle, les pdf sont archivés dans un dossier spécifique. J'ai travaillé sur un corpus de 75 publications au final. La période 1761-1910 n'a pas été prise en compte en raison d'une présentation de données très différentes du format de publication scientifique standard que l'on connaît depuis la moitié du XXème siècle, ces publications annoncent généralement une liste d'espèces découvertes associées à des caractéristiques morphologiques spécifiques.

Période	Nombre de publications
2020-2029	16 fichiers
2010-2019	26 fichiers
2000-2009	10 fichiers
1990-1999	9 fichiers
1960-1989	14 fichiers
1926-1959	2 fichiers

Fig n°17 : Répartition du nombre de publications scientifiques pdf par an dans le cadre de mon projet

Ces publications scientifiques peuvent relater une expérimentation ou faire un état de l'art de type *review*. Un *data paper* est aussi disponible de ce corpus. Les publications sont écrites en majorité en anglais, il existe cependant quelques publications en français. Nicolas Sauvion a notifié dans sa liste de références bibliographiques les noms des auteurs, la source, le type de document (chapitres de livres, articles scientifiques, actes de colloques, HDR, avis d'expertise, etc), l'accessibilité au document (libre accès, sous abonnement, papier, etc), etc.

Les formats d'origine de ces publications peuvent être le format papier ou pdf natif ou html ou xml. J'ai travaillé à partir du format pdf de toutes ces publications référencées, soit de type pdf natif ou pdf image ocrisé (version éditeur numérisée ou OCR avec version payante Adobe Acrobat ou Nuance)

Dans le cadre de mon travail, à partir du corpus de publications scientifiques sélectionnées, l'étape suivante a consisté à transformer ces informations globalement non structurées en données structurées et en connaissances formelles.

2) Text-mining avec l'outil Alvis NLP

1.1) Généralités

L'équipe Bibliome, particulièrement Robert Bossy, a développé la plateforme Alvis NLP pour concevoir des workflows d'extraction d'information et faciliter l'étude de corpus. L'EI consiste à extraire automatiquement des fragments de texte qui désignent les objets ou les concepts, à reconnaître les entités nommées (EN) pertinentes et à les mettre en relation. Comme indiqué

dans les objectifs de mon stage, l'EI que je développe exploite Alvis NLP à l'usage duquel je me suis formée sous la direction de Robert Bossy.

Alvis NLP est un moteur de traitement automatisé de corpus, modulaire, configurable. Cet outil a été écrit en Java, construit avec Maven, la licence est Apache License v2. Il est librement accessible et disponible sur Github à l'adresse <https://github.com/Bibliome/alvisnlp.git> pour installation sur son ordinateur.

```
mkdir backup
ls
git clone https://github.com/Bibliome/alvisnlp.git
cd alvisnlp/
ll
mvn build install
sudo apt install maven
mvn build install
mvn install
ll
./install.sh install
mkdir install
./install.sh install
install/bin/alvisnlp -help

wget https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/english.par.gz
ll
gunzip english.par.gz
ll
bin/tree-tagger
bin/tree-tagger english.par
cd
cd alvisnlp/
cd share
ll
cp default-param-values.xml.template default-param-values.xml
gedit default-param-values.xml &
readlink -e ~/TreeTagger/bin/tree-tagger
cd ..
ll
./install.sh install
cd ../stage-myriam/
ll
hg
history
~/alvisnlp/install/bin/alvisnlp main.plan
~/alvisnlp/install/bin/alvisnlp main.plan
~/alvisnlp/install/bin/alvisnlp-0.11.0-SNAPSHOT main.plan
~/alvisnlp/install/bin/alvisnlp main.plan
history
wget https://filesender.renater.fr/download.php?token=d2fb9ba7-2e20-47f2-add8-b18b7f43052f&files_ids=36954619
ll
rm -f download.php?token=d2fb9ba7-2e20-47f2-add8-b18b7f43052f
mv ~/Téléchargements/resources.zip .
ll
unzip resources.zip
~/alvisnlp/install/bin/alvisnlp main.plan
ll
ll output
less output/entities.csv
wc -l output/entities.csv
ll output/
```

Fig n°18 : Installation locale d'Alvis NLP

Alvis NLP permet de construire des workflows de text-mining à partir d'une bibliothèque de modules de traitements à tous les niveaux de granularité, décrits et disponibles à l'adresse <https://bibliome.github.io/alvisnlp/reference/Module-reference#list-of-all-modules>. La documentation <https://bibliome.github.io/alvisnlp/> permet d'accéder entre autres à la liste des options possibles : <https://bibliome.github.io/alvisnlp/Getting-started> Pour annoter en part-of-speech (POS), l'outil [Treetagger](#) a aussi été installé.

Pour utiliser Alvis NLP, la requête sur le terminal est la suivante :

~/alvisnlp/install/bin/alvisnlp main.plan [OPTIONS]

Le fichier utilisé (par exemple “main.plan”) est un fichier XML spécifiant la séquence de modules et leurs paramètres. Ce fichier permet de centraliser le workflow de traitement automatique de textes choisi afin d’en faciliter la reproductibilité.

Cet outil offre la possibilité d’utiliser de multiples algorithmes tels que la projection de lexiques, l’application de patrons, l’apprentissage automatique. Avec cet outil, la représentation des documents et la visualisation des résultats peuvent en être facilitées.

- List of all modules
- Readers
- Export
- Projectors**
- Pattern-matching
- Mappers
- Machine-learning
- NER
- Segmentation
- Linguistic processing
- Miscellaneous

Pattern-matching

Pattern-matching modules matches user-defined patterns on section contents of annotations from a layer.

Module class	Pattern type	Output
Action	Expressions	Action expressions
CartesianProductTuples	Expressions	Tuples
PatternMatcher	Hearst-like patterns	Annotations and tuples
RegExp	Regular expressions	Annotations
MultiRegExp	Regular expressions	Annotations

Mappers

Mapper modules associate data from a dictionary file. Each mapper class accepts a different format for the dictionary.

Module class	Dictionary parameter	Format
ElementMapper	<code>entries</code>	AlvisNLP data structure
FileMapper	<code>mappingFile</code>	tab-separated text
OBOMapper	<code>oboFiles</code>	OBO

Fig n°19 : Image de [quelques modules d'EI disponibles sous Alvis NLP](#)

1.2) Présentation du script utilisé

A partir du corpus de publications scientifiques, il me fallait dans un premier temps extraire le texte brut, c’est-à-dire la séquence de caractères, issus des formats pdf afin d’obtenir des textes segmentés (séquence de mots (tokens) et de phrases).

Pour mieux analyser les données, il est intéressant d’uniformiser les données (mots, EN, POS-tags, lemmes, ...).

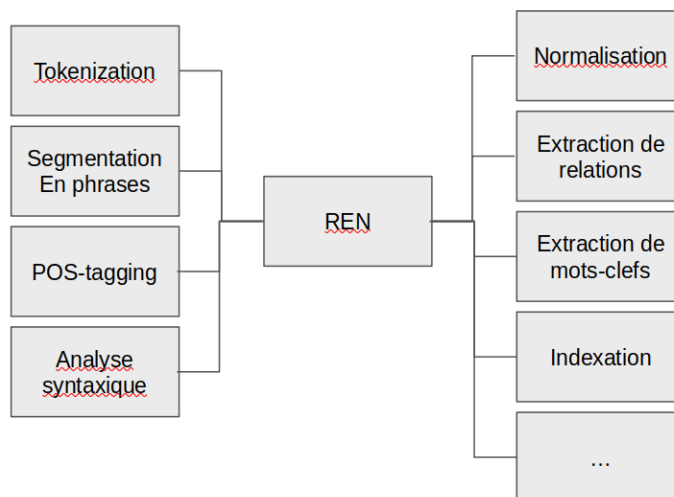


Fig n°20 : Schéma du processus d'EI avec Alvis NLP (d'après Robert Bossy)

Avec les conseils de Robert Bossy, le script Alvis NLP que j'ai fréquemment utilisé dans le cadre de ce stage est composé des modules Alvis NLP décrits ci-après et est détaillé en [ANNEXE n°3](#).

Dans un premier temps, j'ai utilisé le module [TikaReader](#), proposé dans la bibliothèque "[Readers](#)", afin de convertir les **fichiers pdf** en **texte brut**.

Afin d'extraire les EN pertinentes dans le cadre de ce travail, plusieurs fonctionnalités d'Alvis NLP ont été appelées :

- Le script Alvis NLP "taxa.plan" de Robert Bossy a été utilisé afin de désambiguïser l'ensemble des expressions à partir du contexte et de les compiler dans un dictionnaire.

```

<detect href="taxa.plan">
  <taxaDict>resources/taxo/taxa+id_full.txt</taxaDict>
  <compiledDict>resources/taxo/taxa+id_full.trie</compiledDict>
</detect>
  
```

- Les fichiers sous format txt de taxonomie des plantes, des psylles, des phytoplasmes, des lieux géographiques issus du dossier "Ressources" de AlvisNLP ont été utilisés comme dictionnaires. Ces fichiers ont été constitués, notablement enrichis et maintenus à jour par l'équipe Bibliome à partir de la Taxonomie NCBI (Schoch CL *and al.*, 2020). La taxonomie NCBI est mise à disposition en libre accès par le [National Library of Medicine aux Etats-Unis](#). Un des avantages à reprendre cette taxonomie est la classification des données, la nomenclature et la récupération de données synonymes, des noms communs associés aux noms scientifiques.

Ouvrir		taxid_vecteurs.txt -/stage-miriam/resources/taxo	
19165	ncbi:121830	Actizta uncatolides	/ncbi:1/ncbi:131567/ncbi:2759/ncbi:33154/ncbi:33208/ncbi:6072/ncbi:33213/ncbi:33317/ncbi:1206794/ncbi:88770/ncbi:50557/ncbi:85512/ncbi:7496/ncbi:33340/ncbi:33342/ncbi:7524/ncbi:33373/ncbi:33375/ncbi:30092/ncbi:2218240/ncbi:121829/ncbi:121830
19166	ncbi:121831	Blastopsylla	/ncbi:1/ncbi:131567/ncbi:2759/ncbi:33154/ncbi:33208/ncbi:6072/ncbi:33213/ncbi:33317/ncbi:1206794/ncbi:88770/ncbi:60960/ncbi:50557/ncbi:85512/ncbi:7496/ncbi:33340/ncbi:33342/ncbi:7524/ncbi:33373/ncbi:33375/ncbi:1566705/ncbi:121831
19167	ncbi:121832	Blastopsylla occidentalis	/ncbi:1/ncbi:131567/ncbi:2759/ncbi:33154/ncbi:33208/ncbi:6072/ncbi:33213/ncbi:33317/ncbi:1206794/ncbi:88770/ncbi:50557/ncbi:85512/ncbi:7496/ncbi:33340/ncbi:33342/ncbi:7524/ncbi:33373/ncbi:33375/ncbi:1566705/ncbi:121833
19168	ncbi:121833	Boreioglycaspis	/ncbi:1/ncbi:131567/ncbi:2759/ncbi:33154/ncbi:33208/ncbi:6072/ncbi:33213/ncbi:33317/ncbi:1206794/ncbi:88770/ncbi:60960/ncbi:50557/ncbi:85512/ncbi:7496/ncbi:33340/ncbi:33342/ncbi:7524/ncbi:33373/ncbi:33375/ncbi:30092/ncbi:2218198/ncbi:121835
19169	ncbi:121834	Boreioglycaspis melaleucaae	/ncbi:1/ncbi:131567/ncbi:2759/ncbi:33154/ncbi:33208/ncbi:6072/ncbi:33213/ncbi:33317/ncbi:1206794/ncbi:88770/ncbi:50557/ncbi:85512/ncbi:7496/ncbi:33340/ncbi:33342/ncbi:7524/ncbi:33373/ncbi:33375/ncbi:30092/ncbi:2218198/ncbi:121835
19170	ncbi:121835	Cacopsylla	/ncbi:1/ncbi:131567/ncbi:2759/ncbi:33154/ncbi:33208/ncbi:6072/ncbi:33213/ncbi:33317/ncbi:1206794/ncbi:88770/ncbi:60960/ncbi:50557/ncbi:85512/ncbi:7496/ncbi:33340/ncbi:33342/ncbi:7524/ncbi:33373/ncbi:33375/ncbi:30092/ncbi:2218198/ncbi:121835
19171	ncbi:121836	Cacopsylla brunnetipennis	/ncbi:1/ncbi:131567/ncbi:2759/ncbi:33154/ncbi:33208/ncbi:6072/ncbi:33213/ncbi:33317/ncbi:1206794/ncbi:88770/ncbi:50557/ncbi:85512/ncbi:7496/ncbi:33340/ncbi:33342/ncbi:7524/ncbi:33373/ncbi:33375/ncbi:30092/ncbi:2218198/ncbi:121835
19172	ncbi:121837	Cacopsylla myrthi	/ncbi:1/ncbi:131567/ncbi:2759/ncbi:33154/ncbi:33208/ncbi:6072/ncbi:33213/ncbi:33317/ncbi:1206794/ncbi:88770/ncbi:50557/ncbi:85512/ncbi:7496/ncbi:33340/ncbi:33342/ncbi:7524/ncbi:33373/ncbi:33375/ncbi:30092/ncbi:2218198/ncbi:121835/ncbi:121837
19173	ncbi:121838	Cacopsylla peregrina	/ncbi:1/ncbi:131567/ncbi:2759/ncbi:33154/ncbi:33208/ncbi:6072/ncbi:33213/ncbi:33317/ncbi:1206794/ncbi:88770/ncbi:50557/ncbi:85512/ncbi:7496/ncbi:33340/ncbi:33342/ncbi:7524/ncbi:33373/ncbi:33375/ncbi:30092/ncbi:2218198/ncbi:121835/ncbi:121838
19174	ncbi:121839	Cacopsylla pyri	/ncbi:1/ncbi:131567/ncbi:2759/ncbi:33154/ncbi:33208/ncbi:6072/ncbi:33213/ncbi:33317/ncbi:1206794/ncbi:88770/ncbi:60960/ncbi:50557/ncbi:85512/ncbi:7496/ncbi:33340/ncbi:33342/ncbi:7524/ncbi:33373/ncbi:33375/ncbi:30092/ncbi:2218198/ncbi:121835/ncbi:121839

Cacopsylla pyri

Taxonomy ID: 121839 (for references in articles please use NCBI:txid121839)

current name

Cacopsylla pyri (Linnaeus, 1758)

homotypic synonym: **Psylla pyri**

Genbank common name: **pear psyllid**

NCBI BLAST name: **psyllids**

Rank: **species**

Genetic code: [Translation table 1 \(Standard\)](#)

Mitochondrial genetic code: [Translation table 5 \(Invertebrate Mitochondria\)](#)

Other names:

common name(s)

European pear sucker, pear psylla

Fig n°21 : Extraits du fichier de taxonomie "vecteurs.txt" et focus d'une information sur ID:121839 issue de <https://www.ncbi.nlm.nih.gov/taxonomy>

- Les EN de type "Personnes" paraissent importantes à relever pour tester l'identification des auteurs de références bibliographiques cités. Les EN de type "Lieux géographiques" seront utiles dans le cadre d'un futur travail pour mieux déterminer l'association espèce-localisation géographique. Afin d'identifier les EN de Personnes, d'Organisations et de Lieux, j'ai fait appel au module d'Alvis NLP basé sur le classifieur [StanfordNER](#).
- J'ai trouvé pratique de pouvoir les visualiser les résultats sous html avec le module [QuickHTML](#). Cependant, pour travailler plus précisément sur les entités nommées pertinentes, j'ai préféré travailler avec le fichier d'export csv de ces entités, fourni grâce au module [TabularExport](#).

1.3) Expérience avec mes propres dictionnaires

Lors des premières semaines de mon stage, une première étape a consisté à mieux appréhender le corpus et mon objet d'étude. Pour ce faire, j'ai lu un échantillon de publications scientifiques et j'ai listé dans un fichier excel partagé la diversité des nommages de chaque classe citée ci-après, constituant ainsi mon propre dictionnaire :

- Phytoplasme : [ANNEXE n°4 Dictionnaire du phytoplasme considéré et des divers nommages](#)
- Vecteur : [ANNEXE n°5 Dictionnaire des psylles considérés et des divers nommages](#)
- Maladie : [ANNEXE n°6 Dictionnaire sur la maladie et ses divers nommages](#)
- Plante hôte : [ANNEXE n°7 Dictionnaire sur la plante hôte et ses divers nommages](#)
- Indices sur la notion de “transmission”, de modalité à relever : [ANNEXE n°8 Dictionnaire de verbes et marqueurs cités et intéressants dans le cadre de cette étude](#)

Je les ai testé avec Alvis NLP en concevant le script détaillé disponible en [ANNEXE n°9](#). Ces dictionnaires m’ont tout d’abord aidé à caractériser visuellement la position des indices qui m’intéressaient au sein des publications consultées.

Finalement, en raison d’un temps de stage et de rédaction de rapport très limité, je n’ai pas encore eu la possibilité d’explorer pleinement ce travail. Je programme de le faire à partir de fin juin et en juillet.

3) Problématique de conversion du format PDF au texte brut

En parallèle de ce travail précédent, après analyse de mes résultats de script Alvis NLP, j’ai aussi identifié une perte de données textuelles intéressantes pour certaines publications plutôt anciennes. Après diagnostic de ce problème, il me fallait trouver une solution plus efficace pour convertir les publications PDF, en particulier celles numérisées au format OCR, en texte brut.

3.1) TikaReader

Le module [TikaReader](#) m’a permis de convertir les documents au format pdf au format txt. Dans le résultat de cette conversion, suivant la qualité de la numérisation, surtout celle des publications anciennes, j’ai observé :

- une mauvaise reconnaissance de caractère : “u” interprété par “a”, non reconnaissance de caractères diacritiques, etc ;
- des insertions malvenues de séquences de mots représentant des en-têtes de page ou en pieds-de page ; des erreurs de retour à la ligne de lecture de colonnes ;
- des mots coupés par le caractère “-” en raison des césures dans le document d’origine.

D’autres outils payants ont été testés (Acrobat Reader Adobe version payante et Nuance PDF Converter version payante) pour interpréter des scans image. Avec Acrobat Reader Adobe, lors de la conversion au format brut, j’ai rencontré des erreurs d’interprétation de caractères diacritiques. Avec Nuance, j’ai demandé la conversion au format txt unicode.

3.2) Grobid

Classiquement un texte scientifique est structuré par des rubriques du type : titre, résumé, mots-clés, introduction, etc. Afin d'identifier automatiquement ces rubriques, j'ai testé la solution Grobid proposée sur Github par Patrice Lopez, à partir des informations données à l'adresse suivante https://github.com/kermitt2/grobid_client_python.

Ma requête dans le terminal était la suivante :

```
grobid_client --input PDFDIR --output GROBIDDIR processFulltextDocument --consolidate_header --consolidate_citations --include_raw_citations --include_raw_affiliations --verbose --generateIDs --force --segmentSentences --config grobid-config.json
```

Le fichier JSON a été paramétré comme suit, faisant appel au serveur public de Grobid :

```
{
  "grobid_server": "https://kermitt2-grobid.hf.space",
  "batch_size": 1000,
  "sleep_time": 5,
  "timeout": 30,
  "coordinates": [ "persName", "figure", "ref", "biblStruct", "formula", "s" ]
}
```

Dans un premier temps, la moitié des fichiers PDF disponibles dans mon corpus ont été convertis.

Patrice Lopez m'a proposé deux autres solutions pour pallier à ce problème de non-conversion de l'ensemble des fichiers :

- l'utilisation en direct du serveur de démo <https://grobid.science-miner.com>

OU

- la modification de deux paramètres
Baisser fortement --n N concurrency for service usage -> essai avec 1
Augmenter le timeout dans le fichier de configuration JSON -> essai avec 300s au lieu de 30s

Ainsi, j'ai pu traiter environ 98 % des PDF scientifiques. Dans les articles très anciens (exemple avec un article datant de 1926), mal numérisés, le titre et les auteurs sont mal identifiés.

```
<div xmlns="http://www.tel-c.org/ns/1.0" xsl:ids="0vtpMEz"><p xsl:ids="77XnanP"><s xsl:ids="9uG66EC">OBSERVATIONS BIOLOGIQUES SUR PSYLLA.</s><s xsl:ids="PDwtrVv">PYRISUGA <ref type="bibr">{HÉMIPT.}</ref></s><s xsl:ids="d52ba8e">par Frank BROCHET.</s></p></div>
<div xmlns="http://www.tel-c.org/ns/1.0"><head xsl:ids="QPEKM9d">18:1</head><p xsl:ids="sAT4TeV"><s xsl:ids="ruc4yvw">Lorsque, vers la fin du mois d'avril, on examine, sur un poirier particulièrement sur ceux qui sont taillés courts ou qui sont en Pspalter -, les jeunes feuilles en train de se développer, on y observe, assez souvent, de petits insectes allés dont la longueur atteint quatre millimètres environ; parfois, on les trouve accouplés.</s><s xsl:ids="GR6gCqf">Leur corps est de couleur noire -avec, chez certains (les ♀) quelques taches jaunes ou rouges.</s><s xsl:ids="9h0XPRC">Ils sont pourvus d'ailes transparentes, assez grandes et larges, disposées comme un toit sur leur abdomen qu'elles recouvrent en entier, ainsi que ses faces latérales.</s><s xsl:ids="d0EMZzx">Ils ne sont pas craintifs; on peut les observer de près; mais, parfois, ils s'échappent par un brusque saut et disparaissent.</s></p><p xsl:ids="DuDABHf"><s xsl:ids="f8Vh9U">>> Ce sont les Psylles du Poirier (Psylla pyrisugii), insectes de l'ordre des Hémiptères.</s></p><p xsl:ids="BkQ466Y"><s xsl:ids="WAaBDEC">Ils sont pourvus d'un rostre et -ainsi que leur nom l'indiquait- se nourrissent en plantant celui-ci dans les tissus de la plante dont ils sucent la sève.</s><s xsl:ids="SEut4yt">D'où viennent ces insectes?</s><s xsl:ids="nqK622X">On suppose qu'ils ont hiberné, cachés dans les fentes de l'écorce des arbres.</s><s xsl:ids="ZgJPR4p">C'est possible.</s><s xsl:ids="VRE4Qq4">Mais, je crois qu'il est préférable de dire qu'on ignore encore où et comment ils ont passé la mauvaise saison.</s></p><p xsl:ids="E25c0hg"><s xsl:ids="E2moy72">On peut observer ces Psylles allées pendant tout le mois de mai et, durant ce temps, ils pondent continuellement.</s></p><p xsl:ids="08XWV72"><s xsl:ids="WV7114">Les oeufs sont très petits; leur longueur atteint environ 3/10 dt. millimètre; cependant, ils sont visibles à l'oeil nu, parce qu'ils sont réunis en grande quantité et disposés en amas, en général, le long de la nervure médiane, sur les feuilles qui viennent de se développer.</s><s xsl:ids="Mudrtzz">Si l'on prend une de ces feuilles et qu'on l'examine au microscope, on constate que ces oeufs sont en forme de poire et qu'ils sont pourvus latéralement, vers la grosse extrémité, d'un court pédoncule qui est implanté dans les tissus de la feuille ( 1 ).</s></p><p xsl:ids="nbVsmYn"><s xsl:ids="T5q65FH">(1) Certains auteurs ont décrit et ont même figuré ces oeufs, comme ayant, en outre, un long prolongement filiforme à l'extrémité pointue.</s><s xsl:ids="50NHdx">Je n'ai pas constaté cela.</s><s xsl:ids="GH9cFBp"><ref type="bibr">Ann. Soc. ent. Fr., xcv [1926]</ref></s></p><p xsl:ids="EWH5qVn"><s xsl:ids="MUNFFMC">Celle-ci -probablement par suite de la lésion produite par cette implantation -
```

Fig n°22 : Extrait du fichier xml de la publication scientifique de Brochet (1926). Le titre de la publication n'a pas été reconnu et se fond dans le corps du texte. Ce titre se distingue juste par le format "majuscule"

Dans les articles scientifiques plus récents, dont le format scientifique est standard, la partition <titre>, <auteur>, <abstract>, <KW>, <corps du texte> <introduction>, <ref>, etc est propre. Ce format xml permet de récupérer, entre autres, les références de bibliographie (balise <ref>) citées dans le texte ou en bibliographie, ce qui semble très intéressant à exploiter dans le cadre de ce projet.

```
<?xml version="1.0" encoding="UTF-8" ?>
<TEI xmlns="http://www.tei-c.org/ns/1.0"
  xmlns:xsl="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.tei-c.org/ns/1.0 https://raw.githubusercontent.com/kernitt2/grobld/master/grobld-home/schemas/xsd/Grobld.xsd"
  xmlns:xlink="http://www.w3.org/1999/xlink" ?>
  <telHeader xml:lang="en" ?>
    <fileDesc ?>
      <titleStmnt ?>
        <title level="a" type="main" xml:id="EUPE6Kd">Incidence of and Candidatus Liberibacter europaeusand; and phytoplasmas in Cacopsylla species (Hemiptera: Psyllidae)
        and their host/shelter plants</title>
        <funder ref="UKD2X84 #_TEXCS5r">
          <orgName type="Full">unknown</orgName>
        </funder>
      </titleStmnt ?>
      <publicationStmnt ?>
        <publsher />
        <availability status="unknown">
          <licence />
        </availability ?>
        <date type="published" when="2012-03-18">18 March 2012</date>
      </publicationStmnt ?>
      <sourceDesc ?>
        <bibliStruct ?>
          <analytic ?>
            <author ?>
              <persName><forename type="first">Caterina</forename><surname>Camerota</surname></persName>
            </author ?>
            <author ?>
              <persName><forename type="first">Noura</forename><surname>Raddadi</surname></persName>
            </author ?>
            <author ?>
              <persName><forename type="first">Alan</forename><surname>Pizzinat</surname></persName>
            </author ?>
          </analytic ?>
        </bibliStruct ?>
      </sourceDesc ?>
    </fileDesc ?>
  </telHeader ?>
  <text xml:lang="en" ?>
    <div xmlns="http://www.tei-c.org/ns/1.0">
      <head xml:id="Cr7RYnk">Introduction</head>
      <p xml:id="HnEq5A"><span xml:id="eHEVBVK">As vectors of phytopathogenic agents, many psyllids represent important pests of agricultural crops.</span><span xml:id="tRucCZE">The main plant pathogens associated with and transmitted by psyllids are viruses, such as the Potato rosette virus</span><span type="bibr" target="#b3">(Tenorio et al. 2003)</span><span type="bibr" target="#b10">(Hodkinson 2009)</span><span type="bibr" target="#b11">because of the serious economic losses they cause worldwide.</span><span xml:id="pZzpprG">The genus 'Ca.</span><span xml:id="Fec57rM">Liberibacter' includes phloem-limited Gram-negative bacteria belonging to the  $\alpha$ -subdivision of the Proteobacteria</span><span type="bibr" target="#b12">(Jagoueix et al. 1994)</span><span type="bibr" target="#b13">to date, five species have been identified: Huanglongbing (HLB) disease, also known as citrus greening, is associated with 'Ca.</span><span xml:id="bH7G4uA">Liberibacter asiaticus', 'Ca.</span><span xml:id="m6tcjnw">Liberibacter americanus' and 'Ca.</span><span xml:id="vNkQncM">Liberibacter africanus' and affects many citrus cultivars, as well as wild citrus.</span><span xml:id="jnnDZPW">Candidatus Liberibacter asiaticus' is present in Asia and America.</span><span type="bibr" target="#b14">(Gottwald 2010)</span><span type="bibr" target="#b17">Lin 1956;</span><span type="bibr" target="#b20">McLean and Oberholz 1965;</span><span type="bibr" target="#b37">Teixeira et al. 2005)</span><span type="bibr" target="#b26">(Raddadi et al. 2011)</span><span type="bibr" target="#b27">Furthermore, studies of Cleu in C. pyri revealed high infection rates</span><span type="bibr" target="#b28">(Raddadi et al. 2011)</span><span type="bibr" target="#b29">, suggesting that this bacterium has an endosymbiotic relationship with the insect, as suggested for other Liberibacter species</span><span type="bibr" target="#b6">(Duan et al. 2009)</span><span type="bibr" target="#b5">(Carraro et al. 1998)</span><span type="bibr" target="#b5">(Davies et al. 1992)</span><span type="bibr" target="#b14">Jarascul et al. 2003;</span><span type="bibr" target="#b33">Tedeschi and Alna 2004)</span><span type="bibr" target="#b33">Tedeschi and Alna 2004)</span>, possible interactions may have developed between these two microorganisms in both plants and insects.</p>
    </div>
  </text ?>
  <monogr ?>
    <note type="raw"><span type="bibr" target="#b33">Carraro, L., Ferrini, F., Labonne, G., Ernacora, P., & Lot, N. (2004). Seasonal infectivity of Cacopsylla pruni, vector of European stone fruit yellows phytoplasma. Annals of Applied Biology, 144, 191-195.</span></note>
  </monogr ?>
  <bibliStruct xml:id="b1">
    <analytic ?>
      <title level="a" type="main" xml:id="YNDMDMG">Transmission of pear decline using naturally infected Cacopsylla pyri L</title>
      <author ?>
        <persName><forename type="first">L</forename><surname>Carraro</surname></persName>
      </author ?>
      <author ?>
        <persName><forename type="first">N</forename><surname>Loi</surname></persName>
      </author ?>
      <author ?>
        <persName><forename type="first">P</forename><surname>Ernacora</surname></persName>
      </author ?>
      <author ?>
        <persName><forename type="first">A</forename><surname>Gregoris</surname></persName>
      </author ?>
      <author ?>
        <persName><forename type="first">R</forename><surname>Osler</surname></persName>
      </author ?>
    </analytic ?>
    <monogr ?>
      <title level="j" xml:id="p84q6AT">Acta Horticulturae</title>
      <imprint ?>
        <bblScope units="volume">472</bblScope>
        <bblScope units="page">from="665" to="668" />
        <date type="published" when="1998">1998</date>
      </imprint ?>
    </monogr ?>
    <note type="raw"><span type="bibr" target="#b33">Carraro, L., Loi, N., Ernacora, P., Gregoris, A., & Osler, R. (1998). Transmission of pear decline using naturally infected Cacopsylla pyri L. Acta Horticulturae, 472, 665-668.</span></note>
  </bibliStruct ?>

```

Fig n°23 : Extrait du fichier xml de la publication scientifique de [Camerota and al. \(2012\)](#). On observe une structuration du document xml avec identification à l'aide de balises de titre de la publication scientifique, des auteurs, des parties telles que le résumé, l'introduction, etc. On remarque aussi l'identification de référence bibliographique citée avec la balise <ref> dans le corps du texte, ainsi la structure <bibliStruct> détaillant la référence bibliographique complète dans la zone "Bibliographie"

Avec la figure n°24, on observe qu'il est aussi possible d'isoler les DOI des références citées avec la balise de `<ptr target="https://doi.org/10.1007/bf02374564" />`.

```
<biblstruct xml:id="b5">
  <analytic>
    <title level="a" type="main" xml:id="_6vhPjhd">Évolution du complexe parasitaire des psylles du poirier Psylla pyri et Psylla pyrisuga (Homoptera: Psyllidae) en vergers dans le sud-est
    France au cours de la période hivernale, printanière et estivale</title>
    <author>
      <persName<forename type="first">Armand</forename><forename type="middle">E</forename><surname>Lyoussoufi</surname></persName>
    </author>
    <author>
      <persName<forename type="first">A</forename><surname>Rieux</surname></persName>
    </author>
    <author>
      <persName<forename type="first">R</forename></persName>
    </author>
    <idno type="DOI">10.1007/bf02374564</idno>
    <ptr target="https://doi.org/10.1007/bf02374564" />
  </analytic>
  <monogr>
    <title level="j" xml:id="_RNnp4fX">Entomophaga</title>
    <imprint>
      <biblscope unit="volume">36</biblscope>
      <biblscope unit="issue">2</biblscope>
      <biblscope unit="page" from="287" to="294" />
      <date type="published" when="1991">1991</date>
    </imprint>
    <note type="raw_reference">Armand E, Lyoussoufi A, Rieux R. Évolution du complexe parasitaire des psylles du poirier Psylla pyri et Psylla pyrisuga (Homoptera: Psyllidae) en vergers dans le
    sud-est France au cours de la période hivernale, printanière et estivale. Entomophaga. 1991;36(2):287-294. https://doi.org/10.1007/bf02374564</note>
  </monogr>
</biblstruct>
```

Fig n°24 : Extrait du fichier xml de la publication scientifique de [Civolani \(2023\)](#). On observe une structuration du document xml avec identification de DOI associé à chacune des références bibliographiques citées dans la partie Bibliographie de la publication.

La publication de [Weintraub and al. \(2006\)](#) est associée à un “Supplementary Data” disponible depuis le site de l’éditeur avec la mention “Download Table 1”. Cette annexe est importante dans le cadre de notre étude car elle mentionne des informations pertinentes mais sous forme de tableau. Lors de la conversion de ce format PDF avec l’outil Grobid, ce tableau n’a pu être interprété au format xml. De plus, la figure sous forme de tableau disponible dans la publication de [Jarausch and al. \(2019\)](#) au niveau de la partie “Introduction” n’a pu être interprétée lors de la conversion avec l’outil Grobid.

2019_Jarausch_Psyllid vectors_chapter 3.pdf

3.1 Introduction

The jumping plant-lice or psyllids form the well-defined superfamily Psylloidea which belongs to the suborder Sternorrhyncha of the order Hemiptera. About 4,000 species of this superfamily are described worldwide including about 400 species in Europe (Burckhardt 1994; Burckhardt and Ouvrard 2012). All recognised phytoplasma vectors are found in the genus *Cacopsylla* which is part of the subfamily Psyllinae within the family Psyllidae (Weintraub and Beanland 2006; Burckhardt and Ouvrard 2012). An electronic determination key for the most important *Cacopsylla* species found on Rosaceae in Europe is available at the internet by www.psyllidkey.eu (Burckhardt et al. 2008). Table 3.1 gives an introductive overview of the most important phytoplasma diseases and their agents which are

Table 3.1 Phytoplasma diseases of fruit crops, their associated agents, psyllid vectors and vector's host plant

Psyllid species	Phytoplasma	Disease	Host plant
<i>Cacopsylla picta</i>	' <i>Candidatus</i> Phytoplasma mali'	Apple proliferation	<i>Malus x domestica</i>
<i>Cacopsylla melanoneura</i>	' <i>Candidatus</i> Phytoplasma mali'	Apple proliferation	<i>Crataegus</i> spp., <i>Malus x domestica</i>
<i>Cacopsylla pruni</i> A and B ^a	' <i>Candidatus</i> Phytoplasma prunorum'	European stone fruit yellows	<i>Prunus</i> spp.
<i>Cacopsylla pyri</i>	' <i>Candidatus</i> Phytoplasma pyri'	Pear decline	<i>Pyrus</i> spp.
<i>Cacopsylla pyricola</i>	' <i>Candidatus</i> Phytoplasma pyri'	Pear decline	<i>Pyrus</i> spp.
<i>Cacopsylla pyricola</i>	' <i>Candidatus</i> Phytoplasma pyri'	Peach yellow leaf roll (USA)	<i>Prunus persica</i>
<i>Cacopsylla pyrisuga</i> ^b	' <i>Candidatus</i> Phytoplasma pyri'	Pear decline	<i>Pyrus</i> spp.
<i>Cacopsylla chinensis</i>	' <i>Candidatus</i> Phytoplasma pyri' strain PD-TW	Pear decline-Taiwan	<i>Pyrus</i> spp.
<i>Cacopsylla qianli</i> ^b	' <i>Candidatus</i> Phytoplasma pyri' strain PD-TW	Pear decline-Taiwan	<i>Pyrus</i> spp.

^aComplex of two cryptic species A and B
^bPresumed vectors

Fig n°25 : Extrait du fichier xml de la publication scientifique de [Weintraub \(2006\)](#). On observe une figure sous forme de tableau au sein de la partie “Introduction”

Chapitre 3 Vérification de l'information primaire : présentation de deux cas d'étude

Afin de répondre aux objectifs fixés dans le cadre de mon stage ([cf. Introduction, Problématique](#)), pour mieux valider les connaissances scientifiques décrites précédemment ([cf. Chapitre 1](#)), il me semblait très pertinent de partir de cas concrets pour qualifier l'évolution de l'information citée au fil des années. Parmi plusieurs cas concrets, j'ai choisi deux études de cas. Ces deux cas d'étude, l'une axée sur la reprise de l'information de la publication de [Lemoine \(1991\)](#) et, l'autre axée sur la question "l'insecte *Cacopsylla pyrisuga* est-il un vecteur avéré du phytoplasme '*Candidatus Phytoplasma pyri*' ?" sont présentées ci-après.

1) Cas d'étude illustré avec la publication de Lemoine (1991)

Très rapidement, j'ai détecté qu'il fallait travailler sur les zones de texte avec des reprises de citation d'auteurs et tout particulièrement, sur des auteurs de référence souvent cités. A un moment donné, je pensais travailler sur l'étude de citation de la publication de [Jenson \(1964\)](#) mais les circonstances présentées ci-après ont fait que mon cas d'étude s'est plutôt focalisé sur la publication de l'auteur français [Lemoine \(1991\)](#).

1.1) Introduction à ce cas d'étude : Analyse conjointe d'une publication d'une auteure de référence, experte reconnue au niveau international

Durant la période de mon stage, deux stagiaires biologistes travaillaient avec N. Sauvion sur le même sujet (psylles vecteurs) mais sur des questions scientifiques différentes de la mienne. En avril, N. Sauvion nous a proposé un challenge sur deux jours pour nous aider à appréhender rapidement le sujet de notre stage en confrontant nos différentes visions et approches du sujet, en interagissant les uns et les autres, à notre convenance, et avec les moyens que nous aurions choisis (visio, partage en ligne de documents, etc). L'objectif était de rédiger une analyse critique de la fiabilité du contenu de la publication ([Trivellone, 2019](#)).

Le profil Orcid de l'auteure (<https://orcid.org/0000-0003-1415-4097>, information disponible dans le format "xml" de la publication) nous a renseigné sur le nombre de ses publications et sa notoriété par déduction. L'auteure affirmait avoir créé une base de données en ligne complète et mondiale de références sur l'interaction biologique des Hemiptera-Phytoplasma-Plant (HPP). L'auteure avait collecté 2828 publications de recherche sur la période 1895-2018 principalement à partir de la base de données [Web of Sciences](#) et d'une librairie américaine (Funk Agricultural, Consumer, and Environmental Sciences Library at the University of Urbana-Champaign (Illinois, USA)). Le choix de référencement dans sa base de données était basé sur une sélection très stricte de publications prouvant la détection de phytoplasme pathogène dans la plante hôte par des analyses moléculaires ou par des méthodes d'analyse de transmission.

La publication, de type *datapaper*, est en accès libre, facilement trouvable à l'aide de son doi (<https://doi.org/10.3897/BDJ.7.e32910>) et est disponible aux formats html, pdf et xml. Les données brutes ont été déposées aux formats csv ou tsv sur Zenodo (<https://doi.org/10.5281/zenodo.2532738>). Au cours du travail, il s'est avéré que l'information importante sur le phytoplasme spécifique qui nous intéressait ('*Candidatus Phytoplasma pyri*') n'apparaissait pas dans le *datapaper* mais dans les données brutes sous format tabulé tsv ou csv.

Nous nous sommes rendus compte que l'interaction '*Candidatus Phytoplasma pyri*'-*Cacopsylla pyri*-poirier n'était pas mentionnée pour la France, alors que nous nous attendions à ce qu'elle le soit. Après analyse, nous nous sommes rendus compte que peu de publications étaient référencées pour la France dans cette base de données. Seules deux références françaises apparaissent sur la famille des insectes qui nous intéressaient, les *Psyllidae*.

Plus spécifiquement, seule une référence française était citée pour *Cacopsylla pyri* et dans cette publication, cet insecte est associé au phytoplasme de la maladie des pommiers.

record_id	phytoplasma	host_group	host_family	host_species	country	reference_id	full_reference
284. 1850X A		Insect	Psyllidae	Cacopsylla pyri	France	Lemoine 1991	Lemoine, J. 1991. Déperissement du poirier: rôle de <i>Psylla pyri</i> dans sa dissémination. Arboriculture fruitière, 442: 28-32
361. 1850X B		Insect	Psyllidae	Cacopsylla pruni	France	Thebaud et al 2009	Thebaud, G., Vron, M., Alary, R., Sauvion, N., Labonne, G. 2009. Efficient transmission of ' <i>Candidatus Phytoplasma prunorum</i> ' is delayed by eight months

Fig n°26 : Extrait des [données csv](#) avec un focus sur la France et la famille des *Psyllidae* de la publication ([Trivellone, 2019](#))

Cette association nous a intrigués d'autant que N. Sauvion pensait même qu'elle était fausse. Pour vérifier la véracité de cette citation, il nous a fallu rechercher la publication originale de [Lemoine \(1991\)](#).

1.2) Accessibilité au document et à l'information contenue dans la publication de [Lemoine \(1991\)](#)

En France, *Cacopsylla pyri* est considéré comme le vecteur du phytoplasme responsable de la maladie du Pear decline. Il est admis par la communauté des experts biologistes des psylles que l'espèce *C. pyri* n'est pas associée à la transmission d'autres phytoplasmes. Or, Trivellone (2019) affirme clairement une association *C. pyri* avec la transmission de l'Apple Proliferation en citant explicitement la référence ([Lemoine, 1991](#)). Trivellone (2019) sous-entend ainsi que *C. pyri* est vecteur de l'agent pathogène responsable de cette maladie sur pommiers, en l'occurrence '*Candidatus Phytoplasma mali*'.

Pour vérifier la véracité de l'information rapportée par la publication ([Trivellone, 2019](#)), il fallait accéder à la publication ([Lemoine, 1991](#)). Lemoine a publié en 1991 un article intitulé "Déperissement du poirier: rôle de *Psylla pyri* dans sa dissémination" dans la revue "Arboriculture fruitière" destinée aux professionnels agricoles. Récupérer le PDF de cet article ne fut pas une tâche aisée. Rechercher cette publication, issue d'une revue sans renommée scientifique et rangée dans une bibliothèque papier quelque part en France, aurait pu durer plusieurs jours. Mais N. Sauvion a contacté une professionnelle de bibliothèque, Dominique Fournier (SupAgro Montpellier), qui a renvoyé l'article scanné (format PDF) en moins de 2h. En

l'état, l'article donnait l'information rapidement pour un lecteur humain mais n'était pas exploitable par la machine.



Fig n°27 : Lemoine, 1991

- Statut "scientifique" de la revue

Cette publication de Lemoine (1991) n'est pas référencée dans les résultats des deux grandes bases de données scientifiques internationales citées précédemment, Scopus et Web of Science (WOS), car la revue "Arboriculture fruitière" n'est pas une revue à comité de lecture (articles non relus par les pairs).

De même, je me suis intéressée à deux autres articles de ce même auteur publiés en 1997 et 1998 dans Phytoma, une autre revue à destination des professionnels de la filière fruits, sans relecture par les pairs et non référencée dans Scopus ou WOS. L'éditrice actuelle de Phytoma a été contactée par N. Sauvion. Elle a renvoyé quelques jours plus tard les articles scannés. J'ai pu ainsi constater que ces articles de vulgarisation et d'information très générale n'apportaient pas d'informations scientifiques permettant de vérifier la véracité des faits avancés. Ils ne présentaient donc pas d'intérêt pour ma démarche. Je m'attendais au même constat pour l'article Lemoine (1991).

- Travail de très grande qualité scientifique mais non relu par les pairs

Objectivement, sans relecture par les pairs, la fiabilité de l'information contenue dans des articles dans des revues telles que Phytoma ou l'Arboriculture Fruitières peut être questionnée. Or l'article Lemoine (1991) "sort du lot" car il rapporte un travail sérieux, de bonne qualité scientifique, avec un descriptif détaillé des méthodes et des résultats convaincants. Selon N. Sauvion, ce travail aurait même mérité d'apparaître dans une revue scientifique à comité de

lecture. Il lui apparaît donc pertinent et justifié de citer cet article comme étant le premier à démontrer formellement le statut de *C. pyri* comme vecteur du phytoplasme responsable du Pear decline en France.

- Article en français, difficilement interprétable par les non-francophones ?

En complément de sa difficulté d'accès, il est aussi à noter qu'il est écrit en français et son interprétation détaillée paraît difficile pour les non-francophones.

1.3) Connaissance scientifique acquise avec cet article

D'après Lemoine (1991), l'insecte *Psylla pyri* (syn. aujourd'hui de *Cacopsylla pyri*) joue un rôle dans le dépérissement du poirier et dans la dissémination d'un "agent de type mycoplasme (MLO)" (il fallait comprendre "Mycoplasma-like organism" ; aujourd'hui pour les plantes, le concept "phytoplasme" se substitue à celui de MLO). Donc, nous traduisons avec les termes d'aujourd'hui : '*Cacopsylla pyri*' est vecteur du phytoplasme '*Candidatus Phytoplasma pyri*' qui infecte les poiriers cultivés et cause la maladie du Pear decline.

1.4) Analyse des citations de la référence bibliographique ([Lemoine, 1991](#))

Pour analyser les citations de la référence bibliographique, plusieurs outils pouvaient être utilisés comme Alvis NLP, Google scholar, etc. J'ai fait un essai à partir de notre corpus initial et d'une extraction automatique de l'outil Alvis NLP et la reconnaissance du nom de personnes (voir figure n°28). En parallèle, les résultats de citation de Google scholar ont été aussi analysés. Dans le cas présent, nous nous attarderons plutôt sur l'analyse des résultats de citation de Google Scholar présentés ci-après en figure 29. Nous reviendrons ensuite sur l'analyse des résultats d'Alvis NLP avec le deuxième cas d'étude.

	Person		
/home/mdulor/stage-myriam/documents/2001_Carraro_Pear Decline Cacopsylla pyri.pdf	'Person	2057-2064	'Lemoine
/home/mdulor/stage-myriam/documents/2001_Carraro_Pear Decline Cacopsylla pyri.pdf	'Person	15095-15102	'Lemoine
/home/mdulor/stage-myriam/documents/2006_Weintraub_insect vectors of phytoplasma.pdf	'Person	101401-101408	'Lemoine
/home/mdulor/stage-myriam/documents/2008_Ploaie_Mycoplasma Phytoplasma Detectionin Pear with Pear Decline.pdf	'Person	5384-5391	'Lemoine
/home/mdulor/stage-myriam/documents/2008_Ploaie_Mycoplasma Phytoplasma Detectionin Pear with Pear Decline.pdf	'Person	14735-14742	'Lemoine
/home/mdulor/stage-myriam/documents/2009_Hodkinson_JNH_Review Life Cycle Psyllids.pdf	'Person	164981-164990	'Lemoine J
/home/mdulor/stage-myriam/documents/2011_Liu_Phytoplasmas of two 16S rDNA groups are associated with pear decline in Taiwan	'Person	31138-31145	'Lemoine
/home/mdulor/stage-myriam/documents/2014_Jarausch_Review insect vectors Book COST.pdf	'Person	14352-14359	'Lemoine
/home/mdulor/stage-myriam/documents/2014_Jarausch_Review insect vectors Book COST.pdf	'Person	28717-28724	'Lemoine
/home/mdulor/stage-myriam/documents/2019_Jarausch_Psyllid vectors chapter 3.pdf	'Person	32032-32039	'Lemoine
/home/mdulor/stage-myriam/documents/2021_Moreno_Psyllids as major vectors of plant pathogens.pdf	'Person	5115-5122	'Lemoine
/home/mdulor/stage-myriam/documents/2023_Civolani_Diversity, biology, and managemen.pdf	'Person	54854-54861	'Lemoine

Fig n°28 : Extraction automatique du nom de personne "Lemoine" avec l'outil Alvis NLP

Cacopsylla pyri comme vecteur de 'Candidatus Phytoplasma pyri'
(Chaîne de citations générée d'après les citations GoogleScholar, au 5-06-2024)

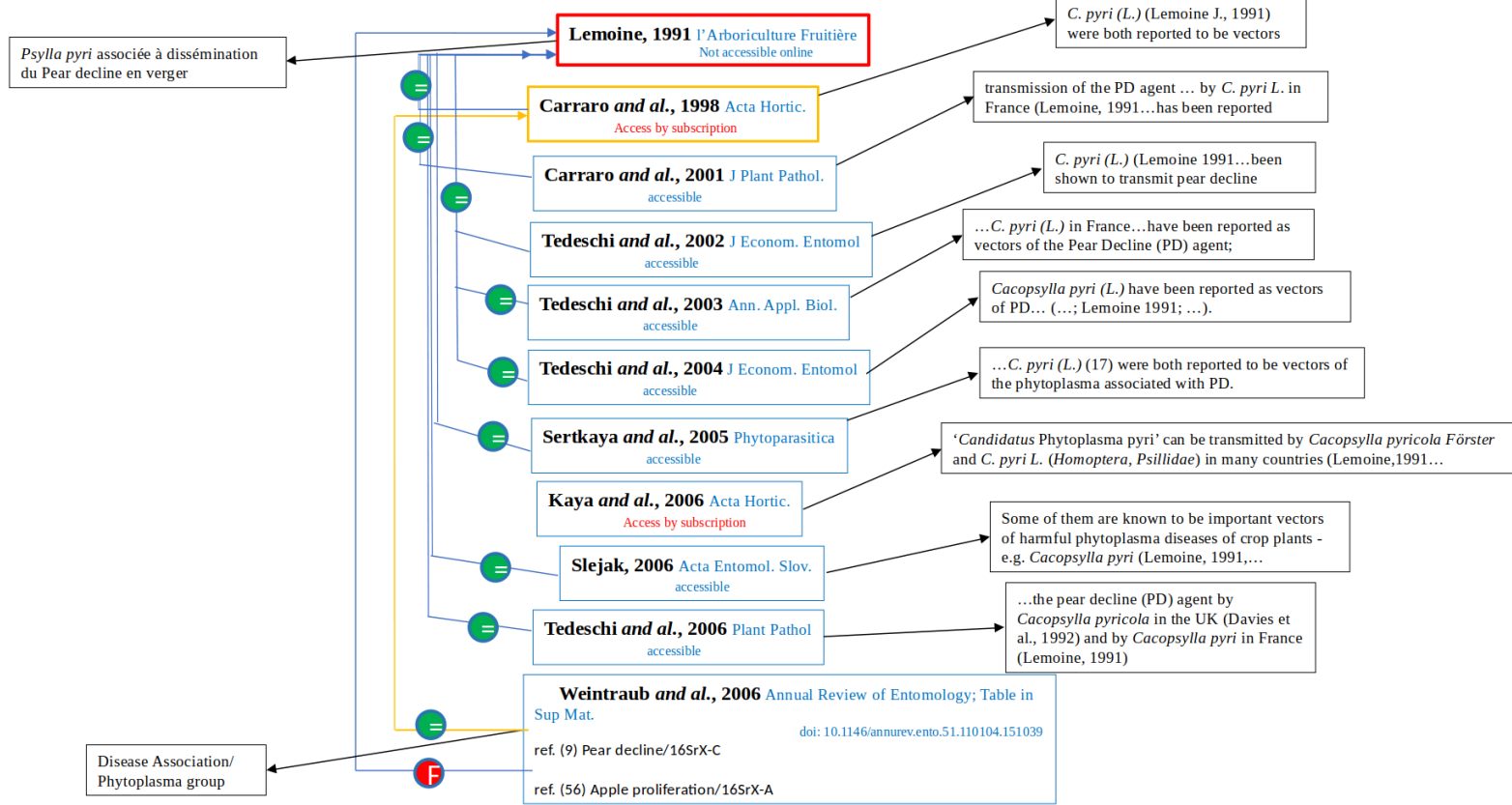


Fig n°29 : Chaîne de citations de (Lemoine, 1991) générée d'après les citations Google Scholar au 5-06-2024.

Analyse focalisée sur l'information : 'C. pyri' comme vecteur de 'Candidatus Phytoplasma pyri' sur la période 1991-2006

La séquence de mots la plus pertinente, extraite de chaque publication, est mise en évidence à droite. Les auteurs de publication sont mentionnés, ainsi que le nom de la revue de la publication considérée et son accessibilité. Les flèches bleues et jaunes symbolisent les citations. La flèche jaune représente l'accès à une publication au format scan image. Les symboles '=' signifie que la citation était équivalente à l'original, 'F' signifie que la citation mentionne des erreurs.

Année publication	nombre citations	PDF	Références (Google Scholar)
1997	97	sub	Carraro, L., Loi, N., Ermacora, P., Gregoris, A., & Osler, R. (1997, June). Transmission of pear decline by using naturally infected <i>Cacopsylla pyri</i> L. In XVII International Symposium Virus and Virus-Like Diseases of T...
1998	223	PDF	Carraro, L., Osler, R., Loi, N., Ermacora, P., & Refatti, E. (1998). Transmission of European stone fruit yellows phytoplasma by <i>Cacopsylla pruni</i> . <i>Journal of Plant Pathology</i> , 233-239.
2000	39	sub	Osler, R., Loi, N., Ermacora, P., Refatti, E., & Carraro, L. (2000, July). Fruit tree phytoplasma diseases diffused in nature by psyllids. In XVIII International Symposium on Virus and Virus-like Diseases of Temperate Fruit...
2001	110	PDF	Carraro, L., Loi, N., & Ermacora, P. (2001). Transmission characteristics of the European stone fruit yellows phytoplasma and its vector <i>Cacopsylla pyri</i> . <i>European Journal of Plant Pathology</i> , 107, 695-700.
2001	53	PDF	Carraro, L., Loi, N., & Ermacora, P. (2001). The life cycle of pear decline phytoplasma in the vector <i>Cacopsylla pyri</i> . <i>Journal of Plant Pathology</i> , 87-90.
2002	139	PDF	Tedeschi, R., Bosco, D., & Alma, A. (2002). Population dynamics of <i>Cacopsylla melanoneura</i> (Homoptera: Psyllidae), a vector of apple proliferation phytoplasma in northwestern Italy. <i>Journal of economic entomology</i> , 95...
2002	9	PDF	Cousin, M. T., & Boudon-Padiou, E. (2002). Phytoplasmes et phytoplasmoses: vecteurs, méthodes de lutte et thèmes de recherche. <i>Cahiers Agricoles</i> , 11(2), 115-126.
2003	50	PDF	Tedeschi, R., Visentin, C., Alma, A., & Bosco, D. (2003). Epidemiology of apple proliferation (AP) in northwestern Italy: evaluation of the frequency of AP-positive psyllids in naturally infected populations of <i>Cacopsylla...</i>
2004	109	PDF	Tedeschi, R., & Alma, A. (2004). Transmission of apple proliferation phytoplasma by <i>Cacopsylla melanoneura</i> (Homoptera: Psyllidae). <i>Journal of economic entomology</i> , 97(1), 8-13.
2005	37	PDF	Sertkaya, G., Martini, M., Ermacora, P., Musetti, R., & Osler, R. (2005). Detection and characterization of phytoplasmas in diseased stone fruits and pear by PCR-RFLP analysis in Turkey. <i>Phytoparasitica</i> , 33, 380-390.
2006	1287	PDF	Weintraub, P. G., & Beanland, L. (2006). Insect vectors of phytoplasmas. <i>Annu. Rev. Entomol.</i> , 51, 91-111.
2006	98	PDF	Tedeschi, R., Ferrato, V., Rossi, J., & Alma, A. (2006). Possible phytoplasma transovarial transmission in the psyllids <i>Cacopsylla melanoneura</i> and <i>Cacopsylla pruni</i> . <i>Plant pathology</i> , 55(1), 18-24.
2006	43	PDF	Sejnak, G. (2006). An overview of the current knowledge of jumping plant-lice of Slovenia (Hemiptera: Psylliodes). <i>Acta entomologica slovenica</i> , 14(1), 11.
2006	0	sub	Kaya, K., Sertkaya, G., & Sertkaya, E. (2006, May). Detection of Pear Decline Disease in Pear and Quince in the Eastern Mediterranean Region of Turkey. In XX International Symposium on Virus and Virus-Like Disease...
2008	5	PDF	Ploaie, P. G., Chireacanu, C., Tatu, M., & Fănu, V. (2008). Mycoplasma (Phytoplasma) detection in pear with pear decline, test plants and psyllids in Romania using dot blot immunoassay method.
2008	0	PDF	MATALLAH, H. (2008). Recherche de maladies à phytoplasmes sur rosacées fruitières dans la région du centre Algérien (Doctoral dissertation, INA).
2009	76	PDF	Mayer, C. J., Jarausch, B., Jarausch, W., Jelkmann, W., Vilcinskis, A., & Gross, J. (2009). <i>Cacopsylla melanoneura</i> has no relevance as vector of apple proliferation in Germany. <i>Phytopathology</i> , 99(6), 729-738.
2009	44	PDF	Tedeschi, R., Lauterer, P., Brusetti, L., Tota, F., & Alma, A. (2009). Composition, abundance and phytoplasma infection in the hawthorn psyllid fauna of northwestern Italy. <i>European Journal of Plant Pathology</i> , 123, 301-...
2011	19	PDF	Shu-Ling, L. I. U., Hsu-Lin, L. I. U., Chang, S. C., & Chan-Pin, L. I. N. (2011). Phytoplasmas of two 16S rDNA groups are associated with pear decline in Taiwan. <i>Botanical Studies</i> .
2015	0	PDF	Emese, T. K., & Intézet, M. A. N. ELTÉRŐ VIRULENCIÁJÚ FITOPLAZMA TORZSEK KÖLCSONHATÁSÁNAK SZEREPE A KERESZTVEDETTSÉG KIÁLLAKULÁSÁBAN. <i>Doktori értekezés</i>
2015	2	PDF	Tibenszkyné Kiss, E. (2015). Eltérő virulenciájú fitoplazma törzsek kölcsönhatásának szerepe a keresztvelettség kialakulásában= Interaction between phytoplasma strains of different virulence and its impact on the develop...
2017	0	PDF	Orsolya, V., Balázs, K., Emese, K., Szilvia, O., & Lajos, J. A. MIT TUDUNK A CA PHYTOPLAZMA PRUNORUM FITOPLAZMA TERJEDESÉRŐL MA ÉS MIT GONDOLUNK UGYANERŐL?.
2021	24	PDF	Moreno, A., Miranda, M. P., & Ferreres, A. (2021). Psyllids as major vectors of plant pathogens. <i>Entomologia Generalis</i> , 41(5).
2022	11	PDF	Raedle-Bauer, M., Paleskić, C., Schönhuber, C., Staples, M., & Brader, G. (2022). Vector transmission and epidemiology of <i>Candidatus Phytoplasma pyri</i> in Austria and identification of <i>Cacopsylla pyrisma</i> as new patho...
2023	12	PDF	Civolani, S., Soroker, V., Cooper, W. R., & Horton, D. R. (2023). Diversity, biology, and management of the pear psyllids: a global look. <i>Annals of the Entomological Society of America</i> , 116(6), 331-357.
2024	0	PDF	Czarnobai De Jorge, B., Kolmann, A., Hummel, H. E., & Gross, J. (2024). Evaluation of a push-and-pull strategy using volatiles of host and non-host plants for the management of pear psyllids in organic farming. <i>Frontiers</i>

Fig n°30 : Liste bibliographique de la chaîne de citations de "Lemoine 1991" générée par Google Scholar au 5-06-2024

L'analyse de la figure n°29, et la chaîne de citation de la connaissance "*Psylla pyri* est associée à une dissémination du Pear decline en verger" de la publication de [Lemoine \(1991\)](#), est détaillée ci-après selon plusieurs axes.

- Analyse du point de vue bibliographique

Que ce soit par analyse manuelle, par l'outil Alvis NLP ou Google Scholar, nous remarquons une reprise très importante de la référence Lemoine (1991). L'accès à la publication [Lemoine \(1991\)](#) n'est pas possible depuis Google Scholar, par contre la citation de cette référence bibliographique est facilement cherchable avec les outils d'aujourd'hui.

D'un point de vue analyse bibliographique, il s'avère que la revue "Annual Review of Entomology" (ARE) est une revue très sélective, donc considérée de très haute qualité dans la communauté scientifique des entomologistes. A priori, les informations contenues dans ce type de revue sont très fiables. En ce sens, il est attendu que toutes les références citées aient été relues soigneusement par le ou les auteurs. La publication de [Weintraub and al. \(2006\)](#) est précisément une review publiée dans ARE

Nous avons remarqué que la publication de Trivellone (2019) évoquée auparavant et essentielle pour notre sujet, n'apparaît pas dans les résultats de Google Scholar. En effet, comme mentionné précédemment l'information qui nous intéressait n'apparaissait pas dans le corps du data paper mais seulement dans les fichiers csv ou tsv déposés dans Zenodo. Google Scholar a seulement extrait les informations au niveau du data paper. L'algorithme de recherche textuel est ainsi passé à côté de cette citation bibliographique issue de l'"annexe" sous format csv/tsv.

- Focus particulier sur la forme textuelle d'un contenant

La publication de [Weintraub and al. \(2006\)](#) fait un état de l'art de la connaissance acquise sur les insectes vecteurs de phytoplasme. L'information sur phytoplasme qui nous intéresse,

'*Candidatus Phytoplasma pyri*', apparaît seulement en annexe de l'article et non pas dans le corps du texte principal de l'article. Nous notons aussi que cette information est relatée sous forme de tableau, structure que je n'ai pas encore réussi à extraire avec les algorithmes d'Alvis NLP ou de l'outil Grobid.

Supplemental Material: Annu.Rev.Entomol. 2006. 51:91-111
doi: 10.1146/annurev.ento.51.110104.151039
Insect Vectors of Phytoplasmas
Weintraub and Beanland, 2006

Vector species	Disease Association/ Phytoplasma group	Host Plants	Distribution
<i>Javesella discolor</i> (Boheman)	Pseudoclassic stolbur (8)	Various families	Europe
<i>Nilaparvata lugens</i> Stål	unnamed phytoplasma (16)	Rice	Asia
<i>Saccharosydne saccharivora</i> (Westwood)	sugarcane yellow leaf (3)	Sugarcane	Cuba
Derbidae			
<i>Proutista moesta</i> (Westwood)	Coconut root wilt/16SrIV (73) Grassy shoot disease (78)	Palms Sugarcane	Southeast Asia Southeast Asia
Flatidae			
<i>Metcalfa pruinosa</i> (Say)	16SrI-B and -G (21)	Various families	Europe
Psyllidae.			
<i>Bactericera trigonica</i> Hodkinson	16SrXII-A (27)	Carrots	Canary Islands
<i>Cacopsylla melanoneura</i> (Forster)	Apple Proliferation/16SrX-A (91)	Apple	Europe
<i>Ca. (=costalis) picta</i> (Forster)	Apple proliferation/16SrX-A (30)	Apple	Europe
<i>Ca. pruni</i> Scopoli	European stone fruit yellows/16SrX-B (11)	Stonefruit trees	Europe
<i>Ca. pyri</i> (Linnaeus)	Apple proliferation/16SrX-A (56) Pear decline/16SrX-C (9)	Apple Pear	Europe Europe


56. Lemoine J. 1991. Deperissement du poirier: role de *Psylla pyri* dans sa dissemination. *Arboric. Fruit* 442:28--32

9. Carraro L, Loi N, Ermacora P, Gregoris A, Osler R. 1998. Transmission of pear decline by using naturally infected *Cacopsylla pyri* L. *Acta Hortic.* 472:665--68

Fig n°31 : Extrait de l'article de publication scientifique (Weintraub and al., 2006)


- Analyse de l'information ayant trait à "*Cacopsylla pyri*" comme vecteur de '*Candidatus Phytoplasma pyri*' à partir de l'analyse des références bibliographiques, des citations (eq. extrait ou reformulation), des informations/découvertes/constatations attribuées à un auteur ou une référence, du contexte de la référence contenant la citation directe ou la reformulation.

Il est à noter que la reprise de cette citation de Lemoine (1991) montre qu'aucun expert ne remet en cause les connaissances scientifiques acquises par Lemoine. Ayant eu nous-même difficilement accès à la publication, la question de fond se pose : les auteurs citants ont-ils réellement eu accès à cette publication et l'ont-ils lue et comprise ?

Le symbole  montre une **reprise sans remise en cause** de la connaissance de Lemoine 1991 : *C. pyri* est le vecteur associé au Pear decline en France. Dans le cas présent, nous

observons dans les encarts d'extraits textuels, la citation de l'auteur Lemoine dans la phrase, souvent encadrée de parenthèses, incluant la date.

Les verbes “report”, “show”, “to be known” dans certains cas apparaissent. Kaya (2006) a mentionné plutôt une possibilité “can be transmitted by”. N. Sauvion confirme que pour lui-même tous ces termes disent la même chose “Lemoine a montré que, le phytoplasme peut être transmis par *C. pyri* selon ...”

Le symbole  montre une **reprise fausse** de la connaissance de l'auteur Lemoine. Cette information n'est pas transcrite dans une phrase mais synthétisée dans un tableau (voir figure n°31). L'information est associée à la référence n° 56 qui renvoie à la référence de Lemoine 1991 ; Dans le tableau, la maladie citée est “Apple proliferation”, au lieu de “Pear decline” ou “dépérissement du poirier” dans le document primaire ; le groupe de phytoplasme associé est “16SrX-A”, au lieu de “16SrX-C” pour le phytoplasme du Pear decline.

Finalement, il apparaît que la publication ([Weintraub and al. , 2006](#)), bien que parue dans une revue prestigieuse, **propage une information fausse en l'attribuant indûment à la publication ([Lemoine, 1991](#))**. A partir de ce diagnostic, nous comprenons comment l'auteure Trivellone (2019) a pu associer par erreur la publication ([Lemoine, 1991](#)) à la notion “ le vecteur *Cacopsylla pyri* transmet le phytoplasme de la maladie des pommiers (“Apple proliferation”)”.

2) Cas d'étude illustré avec l'espèce “*Cacopsylla pyrisuga*”

Afin de poursuivre l'identification d'indices potentiellement importants pour l'extraction automatique d'informations, j'ai aussi mené une enquête sur l'évolution des citations dans le corpus considéré sur l'espèce “*Cacopsylla pyrisuga*” en faisant une EI à partir du mot-clé “pyrisuga”.

J'ai lancé mon script de workflow d'Alvis NLP pour une extraction des EN à partir de la ressource “Taxonomie” de l'équipe Bibliome. L'extraction s'est faite à partir du corpus de pdfs fournis par N. Sauvion. J'ai récupéré un classement des EN sous format csv. J'ai isolé toutes les mentions de l'espèce “*Cacopsylla pyrisuga*” de l'ensemble des espèces citées. Il m'est apparu important de récupérer le contexte, c'est-à-dire les séquences de mots précédents et suivants ce mot-clé “*Cacopsylla pyrisuga*”. Pour l'exemple considéré, j'ai fait une extraction manuelle du contexte. Afin de reproduire cette analyse dans le futur, j'ai inclus cette extraction automatique de contexte avant / après dans le script Alvis LP.



Auteur (1*)	Annee	Mot-clé	Extrait de texte	Citation
			They are shown to comprise seven distinct species: <i>C. pyrisuga</i> (Forster), <i>C. pyrri</i> (L.) (summer form <i>pyraorbis</i> , winter form <i>pyri</i>), <i>C. pyricola</i> (Forster) (summer form <i>pyricola</i> , winter form <i>simulans</i>), <i>C. bidens</i> (Sulc) (summer form <i>vasilevi</i> sylv., winter form <i>bidens</i>), <i>C. notata</i> (Flor) sp. rev., <i>C. permixta</i> sp. n. (= <i>simulans</i> sensu Loginova and Gegechkori nee Forster) and <i>C. fera</i> (Baeva). <i>Psylla horvathi</i> Sulc and <i>P. vicina</i> Sulc are synonymized with <i>C. pyricola</i> . <i>P. picta</i> is regarded as a nomen dubium. Information is given on the host-plants and distribution of each species.	
Burckhardt & Hodkinson	1986	<i>Cacopsylla pyrisuga</i>	Eight species of <i>Pyrus</i> -feeding psyllids, here referred to the genus <i>Cacopsylla</i> Ossiannilsson, were recognized from the Palaearctic Region west of China prior to this present study (Klimaszewski, 1973; Sulc, 1975). Several other species had been sunk in synonymy (Hodkinson, 1984). The recognized species were <i>C. pyricola</i> (Forster), <i>C. pyrri</i> (L.), <i>C. pyrisuga</i> (Forster), <i>C. vasilevi</i> (Sulc), <i>C. vicina</i> (Sulc), <i>C. bidens</i> (Sulc), <i>C. horvathi</i> (Sulc) and <i>C. fera</i> (Baeva). <i>Psylla picta</i> Forster also appears to belong to this group. <i>C. pyricola</i> is recorded as an introduction into both North and South America and, together with <i>C. pyrisuga</i> and <i>C. pyrri</i> , is recorded from the temperate far east. The published geographical distributions and host-plant records of all the above species are summarized by Hodkinson (1984).	KLIMASZEWSKI, S. M. (1973). The jumping plant lice or psyllids (Homoptera, Psyllodea) of the Palaearctic. An annotated check-list.—Annls zool., Warsz. 30, 155-286. (pas de droit de copyright)
Burckhardt & Hodkinson	1986	<i>Cacopsylla pyrisuga</i>	Ossiannilsson (1970) divided the Swedish species of <i>Psylla</i> into subgenera and included <i>pyri</i> , <i>pyricola</i> and <i>pyrisuga</i> in <i>Hepatopsylla</i> . Loginova (1978) subsequently removed <i>pyrisuga</i> to become the type of the new subgenus <i>Thamnosylla</i> and added <i>vasilevi</i> and <i>fera</i> to <i>Hepatopsylla</i> . Klimaszewski (1972), however, suggested that <i>Psylla</i> s.l. should be split into the genera <i>Psylla</i> s.s. and <i>Cacopsylla</i> , including among others the subgenera <i>Hepatopsylla</i> and <i>Thamnosylla</i> . We here formally propose that all the pear-feeding species should be referred to <i>Cacopsylla</i> sensu Klimaszewski.	SULC, K. (1915). Príspevky ku poznani Psyll. III.—Rožp. ceske Akad. (2) 24, 1-35.
Weintraub	2006	<i>Cacopsylla pyrisuga</i>	<i>Cacopsylla pyrisuga</i> (Forster) Pear decline16SfX-C (8) Pears Russia	HODKINSON, I. D. (1984). The taxonomy, distribution and host-plant range of the pearfeeding psyllids.—the publishers are available from the author.
Hodkinson	2009	<i>Cacopsylla pyrisuga</i>		OSSIANNILSSON, F. (1970). Contributions to the knowledge of Swedish psyllids (Hem. Psyllodea) 1-4.—Entomol. Scand. 1, 135-144. https://brill.com/view/journals/isa/1/2/article-p135_9.xml?language=en (via abonnement INRAE)
Hodkinson	2009			LOGINOVA, M. M. (1978). Classification of the psyllid genus <i>Psylla</i> Geoffr. (Homoptera, Psyllidae) [In Russian].—Ent. Obozr. 57, 808-824
Hodkinson	2009			Entomological Review
Hodkinson	2009			Entomol Rev
ARP	2012	<i>Cacopsylla pyrisuga</i>	Les insectes-vecteurs sont le principal moyen de dissémination naturelle de cette maladie (Weintraub & Beanland, 2006). Il existe deux vecteurs essentiellement responsables de la dispersion de 'Ca. P. pyrri': les psylles <i>Cacopsylla pyrri</i> L. et <i>C. pyricola</i> F. (García-Chapá et al., 2005; Seemüller & Harris, 2010). <i>C. pyrri</i> est l'espèce la plus commune en Europe. <i>C. pyricola</i> étant moins fréquent (Siriez, 1982). <i>C. pyrisuga</i> est également mentionnée comme autre insecte-vecteur potentiel de 'Ca. P. pyrri' mais sa transmissibilité n'a pas encore été clairement démontrée (Jarausch & Jarausch, 2010).	ISSN (printed): 0013-8738. ISSN (electronic): 1555-6689.
				KLIMASZEWSKI, S. M. (1972). Bemerkungen über die Systematik der Gattung <i>Psylla</i> Geoffr., s.l. (Homoptera, Psyllodea)—Annls Univ. Mariae Curie-Skłodowska (Sect. C) 27, 11-1. http://bc.umcs.pl/dlibra/pdf/36 . Grbic V. 1974. Some injurious species of the family Psyllidae in pear orchards in Vojvodina. Zastita
				Brocher F. 1926. Observations biologiques sur <i>Psylla pyrisuga</i> (Hemipt.). Ann Soc Entomol Fr. 95:183-188.
				Lazarev MA. 1975. New data on the biology of the pear sucker <i>Psylla pyrisuga</i> (Homoptera: Psyllodea) in the Crimea. Entomol Obozr. 54:758-759.
				Lauterer P. 1999. Results of the investigations on Hemiptera in Moravia, made by the Moravian Museum (Psyllodea 2). Cas Morav Mus., Brno. 84:71-151. (demande par mail le vendredi 17 mai)
				Ossiannilsson F. 1992. The Psyllodea (Homoptera) of Fennoscandia and Denmark. Fauna Entomol Scandin. 26:1-347.
				Jarausch & Jarausch, 2010 : Jarausch B, Jarausch W (2010) Psyllid vectors and their control. In: Weintraub (eds) Phytoplasmas: Genomes, Plants Hosts and Vectors. CAB International, Wallingford, UK, p. 250-270.

Fig n°31 : Extrait de mon fichier d'analyse sous excel. Après avoir extrait avec Alvis NLP les publications contenant l'EN "Cacopsylla pyrisuga" (en bleu), j'ai analysé le contexte textuel (environ 50 mots avant et 50 après) autour de cette EN, en identifiant l'auteur de la référence bibliographique citée (en vert), les citations (eq. extrait ou reformulation), des informations/découvertes/constatations, le contexte de la référence contenant la citation directe ou la reformulation. Je souhaitais par la suite récupérer l'ensemble de ces contextes automatiquement avec Alvis NLP, je n'ai pas eu le temps de finir de travailler là-dessus.

Pour rappel, dans le chapitre 1, j'ai mentionné qu'il restait à prouver comme connaissance scientifique avérée que l'insecte *Cacopsylla pyrisuga* est un vecteur de ce phytoplasme en France et en Europe. La figure n°32, dont l'analyse sera détaillée ci-après, synthétise les observations faites à ce sujet. L'Annexe n°10 complète ces informations.

Les références en noir et en gras représentent les publications scientifiques présentes dans mon corpus d'origine. Les références en gris et en gras symbolisent les publications scientifiques citées importantes à récupérer pour valider la connaissance sur l'espèce *Cacopsylla pyrisuga* (syn. *Psylla pyrisuga*).

Mon analyse a commencé à partir des citations extraites de la publication scientifique (Burckhardt & Hodkinson, 1986).

La séquence de mots la plus pertinente, extraite de chaque publication, est mise en évidence. Les auteurs de publication sont mentionnés, ainsi que le nom de la revue de la publication considérée et son accessibilité. Les flèches bleues ou rouge symbolisent les citations. Les symboles  signifie que la citation était équivalente à l'original,  signifie que la citation mentionne des erreurs, la couleur rouge relève l'erreur.

Les notes en vert relèvent les difficultés d'accès au document original.

* = and al.

Cacopsylla pyrisuga comme vecteur de 'Candidatus Phytoplasma pyri' ? (Exemples de citations)

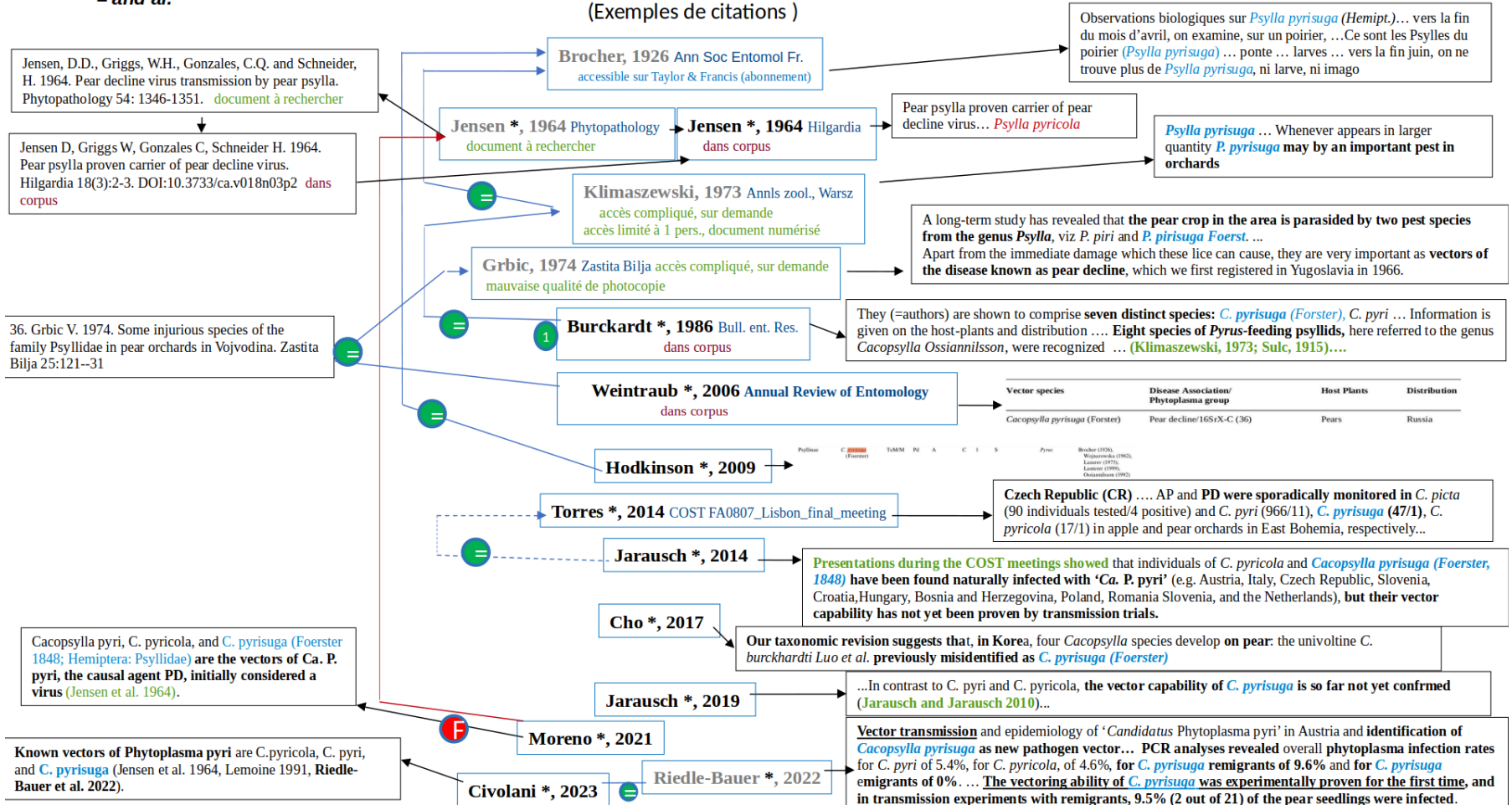


Fig n°32 : Synthèse des résultats des principales citations de *Cacopsylla pyrisuga* (syn. *Psylla pyrisuga*)

2.1) Accessibilité aux documents cités et aux informations contenues dans ces publications

Les publications ([Brocher, 1926](#)), ([Klimasweski, 1973](#)), ([Grbic, 1974](#)), citées respectivement par les auteurs des publications ([Hodkinson and al., 2009](#)) et ([Klimasweski, 1973](#)), par Burckardt and al. (1986) et par Weintraub (2006), n'étaient pas dans le corpus initial.

La publication ([Brocher, 1926](#)) a pu être récupérée à partir d'une recherche de titre entre guillemet dans Google ; le lien renvoyé vers le site éditeur Taylor & Francis ; l'accès a été possible grâce à l'authentification de l'abonnement du service de Documentation de l'INRAE. Le document, même ancien, avait été numérisé par l'éditeur. Le document est écrit en français.

La publication de Klimasweski (1973) a pu être consultée (mais non conservée) grâce au prêt numérique d'un musée. Le document est écrit en anglais.

Après plusieurs jours d'investigation, la publication de Grbic (1974) a pu être récupérée avec des droits de copyright très limités grâce aussi à la bibliothèque d'un musée. L'exemplaire était archivé dans un livre papier relié et une photocopie plutôt de mauvaise qualité a été faite. Le document est écrit en croate.

La publication ([Moreno, 2021](#)) cite la publication ([Jensen, 1964, b](#)). Dans mon corpus d'origine, je disposais d'une publication de Jensen datant de 1964 ([Jensen, 1964 a](#)). En faisant le comparatif de la référence citée et de la référence à disposition, je me suis rendue compte que ce n'était pas la même publication de 1964 citée. Ne disposant pas de celle citée, j'en ai déduit grossièrement que l'information dont je disposais était probablement très similaire à celle que je n'ai pas, puisqu'elles ont été publiées la même année et par le même auteur.

La publication plus récente ([Riedle-Bauer, 2022](#)) a été plus facile à récupérer car elle était citée dans la publication récente ([Civolani, 2023](#)). Aujourd'hui, les auteurs citent dans la bibliographie les références en leur associant les liens et identifiants DOI² quand ils existent ; de ce fait, la consultation des références en est grandement facilitée, dès lors que l'article est en accès libre.

2.2) Connaissance scientifique acquise avec ces articles

Accéder à ces informations primaires était essentiel pour vérifier la vraisemblance des informations publiées dans la littérature scientifique.

1973 : Klimaszewski (document protégé, non diffusable)

"Whenever appears in larger quantity *P. pyrisuga* may be an important pest in orchards"

1974 : Grbic (document compliqué à obtenir + langue croate + mauvaise qualité de photocopie)

Titre traduit en fin d'article : "Some nuisible species of *Psylla* spp., on the pear plantations in Vojvodina"

² DOI = Digital Object Identifiant = identifiant unique d'une publication scientifique

Résumé en anglais disponible à la fin de l'article :

“A long-term study has revealed that **the pear crop in the area is parasided by two pest species from the genus Psylla, viz *P. piri* and *P. pyrisuga* Foerst.** In pear orchards of Vojvodina ***P. piri* is the dominant species, while *P. pyrisuga* occurs only sporadically**, always in association with the former species, and confined to certain localities.

The research has shown that under the given environmental conditions ***P. piri* has five full generations a year, while *P. pyrisuga* develops only one generation on pear and then moves to other hosts.**

Apart from the immediate damage which these lice can cause, **they are very important as vectors of the disease known as Pear decline**, which we first registered in Yugoslavia in 1966.”

2022 Riedle-Bauer

Vector transmission and epidemiology of 'Candidatus Phytoplasma pyri' in Austria and identification of *Cacopsylla pyrisuga* as new pathogen vector

Monika Riedle-Bauer¹ · Caroline Paleskić^{1,2} · Christina Schönhuber² · Martina Staples¹ · Günter Brader² 

“ Pear decline, induced by the phytoplasma '*Candidatus Phytoplasma pyri*', transmitted by pear psyllids, is one of the most devastating diseases on *Pyrus communis* in Europe and North America. Investigations of pear psyllids in 4 pear orchards in lower Austria showed the presence of *Cacopsylla pyri*, *C. pyricola* and *C. pyrisuga* at all locations. PCR analyses revealed overall phytoplasma infection rates for *C. pyri* of 5.4%, for *C. pyricola*, of 4.6%, for *C. pyrisuga* remigrants of 9.6% and **for *C. pyrisuga* emigrants of 0%**. The rates of PCR-positive *C. pyri* and *C. pyricola* individuals varied greatly in the course of the year, and the highest infection rates were observed in late summer, autumn and in late winter. In transmission experiments with healthy pear seedlings, winterform individuals of *C. pyri* and *C. pyricola* transmitted the pathogen to 19.2% (5 out of 26) and 4.8% (2 out of 41) of the test plants, respectively. **The vectoring ability of *C. pyrisuga* was experimentally proven for the first time, and in transmission experiments with remigrants, 9.5% (2 out of 21) of the pear seedlings were infected.**

Fig n°33: Exemples d'extraits "incontournables" pour répondre à la question "l'insecte Cacopsylla pyrisuga est-il un vecteur avéré de ce phytoplasme en France et en Europe ?"

2.3) Analyse des citations

- Focus sur la forme textuelle du contenant

Des exemples de citation sont disponibles en [ANNEXE n°10](#). Il est à noter les différents formats possibles dans le cas présent :

- une phrase avec mention des auteurs à la fin, entre parenthèses ou entre crochets, avant le point final.

“**Eight species of *Pyrus*-feeding psyllids**, here referred to the genus *Cacopsylla* Ossiannilsson, were recognized from the Palaearctic Region west of China prior to this present study (Klimaszewski, 1973; Sulc, 1915).

- un numéro de renvoi à la bibliographie, par exemple (36)

Pear decline/16SrX-C (36)

- un tableau avec les informations à retenir et une liste d’auteurs

Table 1. (Continued.)

Higher taxon	Species	Climate zone	Plant type	Overwintering stage	Overwintering on host	Overwintering elsewhere	Volinism	Feeding site	Gall	Lerp former	Host plant(s)	References
Psyllinae	<i>C. pyrisuga</i> (Foerster)	TeM/M	Pd	A	C	I	S				<i>Pyrus</i>	Brocher (1926), Wojnarowska (1962), Lazarev (1975), Lauterer (1999), Ossiannilsson (1992)

- une phase avec la mention de “reported” ou “have been described by” par exemple

“Seven distinct west Palaearctic *Pyrus*-feeding psyllid species have been described by Burckhardt and Hodinkinson (1986).”


- un raisonnement est à appliquer : dans le cas de deux énumérations distinctes d’espèces et de référence, le lien espèce-référence est donné par le rang dans les énumérations

“Known vectors of *Phytoplasma pyri* are *C. pyricola*, *C. pyri*, and *C. pyrisuga* (Jensen et al. 1964, Lemoine 1991, Riedle-Bauer et al. 2022).”

- Analyse de l’information dans ces reprises de citation : “l’insecte *Cacopsylla pyrisuga* est-il un vecteur avéré de ce phytoplasme en France et en Europe ?”

Dans certains cas, les auteurs des publications du corpus initial citaient la connaissance acquise précédemment par d’autres auteurs sur l’espèce “*Cacopsylla pyrisuga*”. Dans la plupart des cas considérés, ces **citations** étaient **relativement proches** du document primaire. Mais la plupart des informations étaient juste de l’ordre taxonomique sans répondre à la question de vecteur avéré ou pas. Autrement dit, certaines informations ne nous ont pas été utiles pour répondre précisément à la question du statut de vecteur de *C. pyrisuga*, car elles ne mentionnent aucune information sémantique de type “transmet le phytoplasme à la plante hôte”, “est porteur du phytoplasme”.

Il est à noter une évolution du vocabulaire de nommage au fil des années (exemple : “*Psylla pyrisuga*” puis “*Cacopsylla pyrisuga*”, “*Pirus communis*” / “*Pyrus communis*” suivant les langues, etc). Au fil de l’analyse, une évolution de la connaissance scientifique apparaît clairement, il semble nécessaire de construire un référentiel représentant l’apparition de nouveaux concepts (analyse morphologique, moléculaire, génétique, entre autres).

Il est à relever qu'une **citation faussée** (symbolisée par ) a été mentionnée (voir flèche en rouge dans la figure n°32). En effet, Jensen (1964) a fait une étude expérimentale sur l'espèce *Psylla pyricola* et non sur l'espèce *Psylla pyrisuga* comme affirmé par l'auteur de la publication (Moreno, 2021).

"*Cacopsylla pyri*, *C. pyricola*, and *C. pyrisuga* (Foerster 1848; Hemiptera: Psyllidae) are the vectors of 'Ca. P. pyri', the causal agent PD, initially considered a virus (Jensen et al. 1964)."

Dans la publication (Klimaszewski, 1973), j'ai noté l'emploi du verbe modal "may". La **notion de "pest"** prête aussi à confusion : s'agit-il de vecteur ou de ravageur ? dans certains cas, il fait référence à des parasitoïdes (hors-sujet par rapport à l'étude du phytoplasme). Ici, la notion de "pest" est difficilement interprétable sans contexte.

Klimaszewski, 1973 : "Whenever appears in larger quantity *P. pyrisuga* may be an important pest in orchards"

Par rebond de citation, j'ai retrouvé **plusieurs mentions confirmant le rôle de vecteur** de cette espèce, listées ci-après :

- la publication ([Weintraub and al., 2006](#))

Les publications issues de la revue "Annu. Rev. Entomol" font référence dans le cercle des experts. Par conséquent, comme déjà mentionné dans l'analyse du cas concret précédent, la publication de Weintraub (2006) était importante. Elle affirme le fait suivant : selon la publication (Grbic, 1974), le vecteur *C. pyrisuga* transmet le phytoplasme 16SrX-C (eq. 'Candidatus Phytoplasma pyri') au poirier ; ce phytoplasme cause la maladie Pear decline.

Vector species	Disease Association/ Phytoplasma group	Host Plants	Distribution
<i>Cacopsylla pyrisuga</i> (Forster)	Pear decline/16SrX-C (36)	Pears	Russia

36. Grbic V. 1974. Some injurious species of the family Psyllidae in pear orchards in Vojvodina. Zastita Bilja 25:121--31

La publication n° 36 = Grbic (1974) a été publiée en Russie à l'époque de l'ex-URSS ; ce document est très difficilement accessible. Ce fait était douteux sans accès au document primaire.

- la publication ([Grbic, 1974](#))

"Some nuisible species of *Psylla* spp., on the pear plantations in Vojvodina"

"A long-term study has revealed that the pear crop in the area is parasided by two pest species from the genus *Psylla*, viz *P. piri* and *P. pyrisuga* Foerst. In pear orchards of Vojvodina *P. piri* is the dominant species, while *P. pyrisuga* occurs only sporadically, always in association with the former species, and confined to certain localities."

Les notions de "nuisible", de "parasided", de "pest" pourraient prêter à confusion car ces termes aujourd'hui ne font pas références à la notion de vecteur. Ici, grâce au contexte, on comprend

que ces notions utilisées en 1974 font bien référence aux espèces qui nous intéressent, même si les termes de vecteur et bactérie n'apparaissent pas.

- la publication ([Torres and al., 2014](#)) issue d'un groupe de travail européen COST

“**Czech Republic (CR)** AP and PD were sporadically monitored in *C. picta* (90 individuals tested/4 positive) and *C. pyri* (966/11), *C. pyrisuga* (47/1), *C. pyricola* (17/1) in apple and pear orchards in East Bohemia, respectively. Examination of hemipterans in Czech vineyards revealed presence of phytoplasmas belonging to subgroups 16SrI-B, 16SrI-C, 16SrI-F, 16SrIII-B, 16SrXII-A and, unexpectedly, in *Jassargus obtusivalvis* of phytoplasmas of the 16SrXIV group.”

Dans la publication (Torres and al., 2014), grâce à la mention de “(47/1)” accolée à l'espèce qui nous intéresse, on comprend qu'il y a une seule contamination sur 47 donc que le taux de contamination est très faible. Il faudra vérifier si cette information peut être détectée et interprétée automatiquement.

- la publication ([Riedle-Bauer and al., 2022](#))

(Titre) “**Vector transmission** and epidemiology of ‘Candidatus Phytoplasma pyri’ in Austria and **identification of *Cacopsylla pyrisuga* as new pathogen vector**”

(Résumé) “ **Pear decline**, induced by the phytoplasma '**Candidatus Phytoplasma pyri**', transmitted by pear psyllids, is one of the most devastating diseases on *Pyrus communis* in Europe and North America. Investigations of pear psyllids in 4 pear orchards in lower Austria showed the presence of *Cacopsylla pyri*, *C. pyricola* and *C. pyrisuga* at all locations. PCR analyses revealed overall phytoplasma infection rates for *C. pyri* of 5.4%, for *C. pyricola*, of 4.6%, for *C. pyrisuga* remigrants of 9.6% and **for *C. pyrisuga* emigrants of 0%**. The rates of PCR-positive *C. pyri* and *C. pyricola* individuals varied greatly in the course of the year, and the highest infection rates were observed in late summer, autumn and in late winter. In **transmission experiments** with healthy pear seedlings, winterform individuals of *C. pyri* and *C. pyricola* transmitted the pathogen to 19.2% (5 out of 26) and 4.8% (2 out of 41) of the test plants, respectively. **The vectoring ability of *C. pyrisuga* was experimentally proven for the first time, and in transmission experiments with remigrants, 9.5% (2 out of 21) of the pear seedlings were infected.** Our data indicate a significant risk of pathogen transmission in pear orchards during the greater part of the year, especially in late winter, early spring and autumn. Multilocus sequence analysis by aid of the genes *aceF* and *imp* allowed the discrimination between 15 phytoplasma types. Three so far undescribed *aceF* genotypes and four undescribed *imp* genotypes were identified.”

Dans cette publication (Riedle-Bauer and al., 2022), relatant une expérimentation, il est annoncé dès le titre les expressions “vector transmission”, “identification ... as new pathogen vector”. Les expressions “transmission experiments” et “the vectoring ability of *C. pyrisuga* was experimentally proven for the first time” sont aussi des informations très importantes à relever dans le cadre de notre étude et de son automatisation. Une nouvelle notion est ajoutée, la notion d'espèce “remigrant” ou “emigrant”.

Dans d'autres cas, en particulier dans des revues d'état de l'art, l'**affirmation est modérée** :

- la publication ([Jarausch and al., 2019](#))

“In contrast to *C. pyri* and *C. pyricola*, the vector capability of *C. pyrisuga* is so far not yet confirmed (Jarausch and Jarausch 2010)”

Le cas ci-après basé sur une expérience morphologique et génétique affirme qu’il y a eu confusion d’espèce, elle n’informe pas toutefois sur le rôle de vecteur :

- la publication ([Cho and al., 2017](#))

“Our taxonomic revision suggests that, in Korea, **four *Cacopsylla* species develop on pear**: the univoltine *C. burckhardti* Luo et al. previously **misidentified as *C. pyrisuga* (Foerster)**”

On peut aussi remarquer que les indices essentiels peuvent se trouver dans le corps du texte ou en annexe ou bibliographie. Le titre et le résumé aident à contextualiser le sujet central de la publication (revue ou expérience primaire, thème généralisé aux Psylles ou portant sur l’étude d’une espèce spécifiquement). Si le thème est par exemple généralisé aux Psylles, les indices essentiels seront à rechercher dans le corps du texte (recherche du mot-clé puis étude du contexte (approximativement 50 mots avant et après)).

Chapitre 4 Discussion & Conclusion

Pour qualifier la vraisemblance et l'originalité des informations publiées dans la littérature scientifique, au fil des expériences décrites dans [le chapitre 3](#), plusieurs indices et diverses problématiques ont émergé. Dans ce chapitre 4, j'ai diagnostiqué et synthétisé quelques-unes des **solutions à apporter à cette recherche de la validité des informations scientifiques publiées**, sur la base de cette expérience de deux mois de stage.

Synthèse concernant l'accès aux informations primaires citées

Pour mesurer le niveau de véracité, négation, spéculation, incertitude, l'importance d'accéder aux informations primaires citées (*eq.* "revenir à la source d'information") a été soulignée afin de les comparer aux informations secondaires. Plusieurs difficultés ont été rencontrées et sont recensées ci-après.

Pour récupérer l'information, la **conversion du format pdf au format texte** n'est pas triviale. Lors de cette conversion, plusieurs déformations ont pu être observées.

S'il s'agit de **format pdf natif**, ce qui est généralement le cas des publications à partir des années 2000, l'information est préservée. Mais encore faut-il disposer d'un **accès à ces publications**. L'accès en sera facilité si la publication est **en libre accès** et si le **lien hypertexte DOI** est mentionné dans la référence bibliographique citée en bibliographie ou dans la liste de références bibliographiques du corpus considéré. L'accès à ce document pdf natif sera moins aisé s'il faut disposer d'un **abonnement à la revue** consultée ou d'une **adhésion à une société savante** pour disposer de contenu en ligne. Si c'est le cas, l'**authentification automatique** pour accéder à ces publications à accès restreint est plus pratique pour un téléchargement en masse.

S'il s'agit d'un **format pdf numérisé avec OCR**, ce qui est majoritairement le cas pour les publications avant les années 2000, des **erreurs d'interprétations des caractères** peuvent être observées suivant la **qualité de l'océrisation**. Ces erreurs dues à la conversion peuvent être conséquentes si elles tronquent les termes importants dans le cadre de l'étude comme par exemple "Psynllu pyricolu" au lieu de "Psynlla pyricola". J'ai aussi remarqué des variations orthographiques, présentes à l'origine, suivant l'époque ou la langue de la publication, comme par exemple "Pirus communis" au lieu de "Pyrus communis", "Psynlla pricola" au lieu de "Psynlla pyricola". Peut-être faudrait-il inclure dans un processus d'extraction d'information (EI) automatique, le calcul de similarité de mots et la validation ou l'acceptation d'un taux d'erreur de deux caractères peut-être. Bien entendu, la meilleure qualité de numérisation devra être recherchée. J'ai pu voir des numérisations de qualité fournies par les éditeurs eux-même qui présentaient toutefois de légères erreurs d'interprétation de caractère au niveau d'entités nommées (EN) importantes, il semble donc important de réfléchir à la solution à apporter à ce problème. Ce pourrait être aussi un dictionnaire recensant toutes ces variantes orthographiques. Ou à l'heure des réseaux de neurones, l'acceptation d'une variation en fonction des poids et des contextes calculés.

Les publications éditées au format papier dans de gros livres reliés sont plus difficiles à numériser. Globalement, **les articles parus avant les années 1955 sont peut-être moins faciles à numériser.**

Au cours des expériences faites, comme je l'ai souligné dans [le chapitre 3](#), il était primordial d'avoir accès à l'ensemble des publications de référence sur le sujet identifiées à partir du corpus de fichiers pdfs à analyser. Pour le travail futur d'EI automatique sur le sujet, la situation idéale serait de travailler sur un **corpus de publications pdf le plus exhaustif possible**. Cela revient à faire une recherche d'information (RI) la plus exhaustive possible. Les cas d'étude avec par exemple l'article de Lemoine (1991) ou avec l'article de Grbic (1974) ont mis en exergue l'importance d'accéder à ces **précieux documents "rares"** ou la "**littérature grise**³". J'ai déjà mentionné que cela n'était pas une tâche facile, même pour les professionnels de documentation avertis et qu'à ce jour, il n'existait aucune base de données exhaustive sur un sujet donné.

Une solution pour aller vers une connaissance exhaustive serait d'utiliser plusieurs outils pour accéder à un ensemble d'informations plus riche en combinant par exemple des recherches sur Google, Scopus, Web of Science, des bases de données spécialisées dans une discipline, etc. Mais, le peu d'articles obtenus avec les bases de données WOS et Scopus ([cf. chapitre 2. 1](#))), pour les années antérieures à 1990 et tout particulièrement avant les années 1960, démontre une absence d'indexation et de référencement général de données anciennes dans ces bases de données. Par conséquent, pour obtenir des **références antérieures à 1960** (éventuellement à 1990), mieux vaut chercher ailleurs. Au cours de mes recherches d'information, j'ai surtout fonctionné par **rebond de citation issue des bibliographies** de publication. Pour un meilleur résultat, il apparaît important d'arriver en TAL à **extraire les références bibliographiques citées** dans le corps du texte de la publication et dans la partie bibliographie. L'expérience initiée avec l'outil Grobid et la **conversion au format xml** me semble très intéressante à approfondir. J'ai pu constater que Grobid permettait de récupérer les références bibliographiques et en particulier d'extraire les liens hypertextes de DOI ([cf chapitre 2. 3.2](#))) avec l'exemple de la *review* de [Civolani and al. \(2023\)](#). Cette revue de qualité recense beaucoup de publications très pertinentes sur le sujet qui nous intéresse, en y associant les DOI existants dans les références de la bibliographie. Une des stratégies à essayer pour enrichir le corpus en publications de référence serait de partir de cette bonne revue récente pour extraire les références bibliographiques citées à l'aide des DOI. Toutes ces références issues d'une bibliographie ne portent pas sur le sujet qui nous intéresse ; il serait préférable d'extraire les références citées à proximité des EN qui nous intéressent (contexte d'environ 50 mots avant et après). Ce référencement pourrait servir de base (*eq. référentiel des références "pivot"*) puis être enrichi avec l'extraction, moins aisée, des références citées et complémentaires issues de publications moins récentes de proche en proche. Ce type de recensement des références pertinentes se ferait sur le **principe de la comparaison** par

³ Littérature grise selon [Wikipedia en juin 2024](#) : terme générique, désigne les documents produits par l'administration, l'industrie, l'enseignement supérieur et la recherche, les services, les ONG, les associations, etc., qui n'entrent pas dans les circuits habituels d'édition et de distribution.

rapport à ce qui a déjà été indexé dans le référentiel des références “pivot”. J’ai aussi relevé au cours de mes expériences, l’importance de **comparer le titre de la référence complète citée en bibliographie avec le titre de la publication disponible dans le corpus pour un même auteur**. Dans mon cas d’étude, les auteurs citaient fréquemment la publication (Jensen, 1964). Depuis le début de mon stage, une publication ([Jensen and al., 1964 a](#)) était enregistrée dans mon corpus. J’ai toujours pensé en voyant les citations de Jensen *and al.* (1964) des autres auteurs, qu’ils faisaient référence à cette publication déjà enregistrée. Je me suis rendu compte dernièrement, en comparant le titre de la publication, que ce n’était pas la même référence.

Jensen, D.D., Griggs, W.H., Gonzales, C.Q. and Schneider, H. 1964. Pear decline virus transmission by pear psylla. *Phytopathology* 54: 1346-1351. [document à rechercher](#)



Jensen D, Griggs W, Gonzales C, Schneider H. 1964. Pear psylla proven carrier of pear decline virus. *Hilgardia* 18(3):2-3. DOI:10.3733/ca.v018n03p2 **dans corpus**

(cf. [Chapitre 3, Cas d’étude C. pyrisuga](#))

Lorsque les références de documents “rares” ont été identifiées, un autre défi est d’obtenir l’article souhaité. Au fil de mon stage, j’ai pu constater l’importance de travailler en réseau d’experts, aussi bien en entomologie qu’en bibliothèque pour tenter de récupérer “le fameux document”. Après discussion avec Nicolas Sauvion, des érosions en termes d’experts entomologistes très qualifiés sont constatées et ces pertes de “savoir” paraissent inquiétantes. Les départs en retraite ne sont pas remplacés. Je peux dire la même chose en ce qui concerne le service de fourniture de publication à l’unité car ce service considéré “avec peu de valeur-ajoutée” disparaît au sein des organismes. Les bibliothèques ou les archives possèdent des trésors de connaissance rangés dans des rayons de bibliothèque sous format papier. Encore faut-il pouvoir les identifier depuis le web pour y accéder. Un grand **travail de numérisation** de ces documents précieux au service du TAL paraît nécessaire, tout en respectant les **droits d’auteur**.

Avec le *datapaper* de Trivellone (2019), j’ai pu noter qu’une expérience de type **science ouverte** a été initiée. Valeria Trivellone souhaitait mettre en place une grande base de données de référence sur l’interaction biologique des HPP, à enrichir en mode collaboratif. Avec son dépôt de données brutes sous Zenodo, un référentiel de références bibliographiques sur le sujet a été mis à disposition. Certes, dans son format actuel, il ne permet pas d’accéder à l’ensemble des publications citées ([cf. illustration dans chapitre 3, analyse conjointe](#)). Il faudrait au moins rajouter, si possible, le lien hypertexte du DOI aux références bibliographiques complètes qu’elle cite dans son fichier csv ou tsv.

L’idéal serait de disposer d’une **grande base de données** de ces publications avec des **données interopérables**, tout en respectant les droits d’auteur pour effectuer la fouille de texte de manière efficace. Un rêve peut-être ! En se projetant dans le **futur**, il faudrait aussi imaginer l’apparition de nouveaux formats de publication scientifique, comme par exemple la vidéo scientifique.

Accéder à l'information primaire, c'est aussi pouvoir interpréter le document dans n'importe quelle langue. Le sujet de la traduction n'a pas été investigué durant mon stage. Il est à noter que Nicolas Sauvion a souvent recours à l'**outil de traduction** DeepL pour mieux et plus vite comprendre les articles qui l'intéressent, en particulier ceux écrits dans une langue autre que le français ou l'anglais.

L'information primaire citée dans l'information secondaire ne fait pas forcément référence à une phrase bien localisée au sein de la publication contenant l'information primaire. Dans certains cas, les mots-clés de l'information primaire cités sont répartis dans de petites zones textuelles, très espacées les unes des autres au sein de la publication. Des exemples peuvent être illustrés avec les publications relatant des expérimentations comme celles de [Jensen \(1964\) \(a\)](#) ou de [Lemoine \(1991\)](#). Pour mieux comparer l'information primaire de l'information secondaire et les potentielles divergences, faut-il envisager une "fiche de synthèse de connaissance" des publications "pivot" ?

Synthèse concernant les indices bibliographiques, biologiques & linguistiques

Pour mesurer le niveau de véracité, négation, spéculation, incertitude, à l'issue de deux mois de stage, j'ai relevé des **pistes d'indices** qui nécessiteraient des investigations plus approfondies. J'imagine construire peut-être en juillet une **grille des critères d'évaluation de ces indices**. Malheureusement, j'ai eu aussi peu temps pour faire un **état de l'art** conséquent sur le sujet. J'ai relevé des publications intéressantes mais je n'ai pas eu le temps de les analyser.

Indices bibliographiques

Quelques-uns des indices bibliographiques qui seront à prendre en compte dans une grille d'évaluation de critères sont décrits ci-après.

Il faudra tenir compte de la **notoriété de la revue scientifique**, donc le nom de la revue de publication devra être identifié pour faciliter la recherche à ce sujet. Pour évaluer la notoriété de la revue, le facteur d'impact, la mesure des **taux de citation**, la **relecture par les pairs** peuvent être pris en compte. Mais de par expérience terrain, l'**avis d'un expert sénior de la thématique considérée** sur la qualité d'une revue donnée est un point à majorer. De plus, avec le cas de la publication de (Lemoine, 1991), un travail scientifique de très grande qualité peut ne pas être référencé dans une revue scientifique relue par les pairs. Par conséquent, à partir de corpus de références données sur la thématique considérée, un des indicateurs très important sera de mesurer le taux de citation de la référence bibliographique considérée au sein même du corpus.

Il faudra aussi tenir compte du **niveau d'expertise de l'auteur** d'une publication. Avec l'exemple de [Trivellone \(2019\)](#), des indices peuvent être retrouvés sous ORCID ainsi que sur des bases de données d'analyse de publications scientifiques comme SCIVAL, INCITES, etc. Il peut y avoir un ou plusieurs auteurs de publication.

Avec la conversion en XML réalisée par l'**outil Grobid**, j'ai noté la possibilité d'identifier le titre de la publication ainsi que celui des auteurs, en particulier le premier auteur, avec les balises <biblStruct>, <author> <persName> <forename type="first">), ou <title> <monogr><...ISSN>,

exemple en particulier avec la publication. Mais ce n'est pas possible dans tous les cas, exemple en particulier avec la publication ([Weintraub and al., 2006](#)).

```
<sourceDesc>
  <biblStruct>
    <analytic>
      <author>
        <persName><forename type="first">Geonho</forename><surname>Cho</surname></persName>
      </author>
      <author>
        <persName><forename type="first">Daniel</forename><surname>Burckhardt</surname></persName>
      </author>
      <author>
        <persName><forename type="first">Hiromitsu</forename><surname>Inoue</surname></persName>
      </author>
    </analytic>
    <monogr>
      <idno type="ISSN">1175-5326 (</idno>
      <imprint>
        <date type="published" when="2017-12-04">4 Dec. 2017</date>
      </imprint>
    </monogr>
    <idno type="MD5">B3E5D58C486E2CB15E03AD8639866E32</idno>
    <idno type="DOI">10.11646/zootaxa.4362.1.4</idno>
    <note type="submission">Accepted by D. Ouvrard: 18 Oct. 2017;</note>
  </biblStruct>
</sourceDesc>
```

Fig n°33 : Extrait de résultat avec l'outil Grobid, exemple avec la publication ([Cho and al., 2017](#))

```
<sourceDesc>
  <biblStruct>
    <analytic>
      <author>
        <persName><forename type="first">Phyllis</forename><forename type="middle">G</forename><surname>Weintraub</surname></persName>
        <email>phyllisw@volcanti.agri.gov.il</email>
      </author>
      <author>
        <persName><forename type="first">LeAnn</forename><surname>Beanland</surname></persName>
        <email>lbeanlan@vt.edu</email>
      </author>
      <author>
        <persName><forename type="first">Alson</forename><forename type="middle">H</forename><surname>Smith</surname><genName>Jr</genName></persName>
      </author>
    </analytic>
    <monogr>
      <title level="a" type="main" xml:id="_N8zh5XR">INSECT VECTORS OF PHYTOPLASMAS</title>
      <imprint>
        <date type="published" when="2005-09-22">September 22, 2005</date>
      </imprint>
    </monogr>
    <idno type="MD5">D3F6028982C666C144904FF084C5D9C9</idno>
    <idno type="DOI">10.1146/annurev.ento.51.110104.151039</idno>
  </biblStruct>
</sourceDesc>
```

Fig n°34 : Extrait de résultat avec l'outil Grobid, exemple avec la publication ([Weintraub, 2006](#))

Nous avons observé plusieurs catégories de publication comme des chapitres de livre, articles scientifiques, actes de colloque, manuscrit de HDR, avis d'expertise, etc. Les chapitres de livre correspondent généralement à une *review* de la connaissance ; les actes de colloque relatent une expérimentation ; les articles scientifiques correspondent soit à une *review* de la connaissance, soit à une expérimentation. Ce critère bibliographique "review de la connaissance" ou "expérimentation" sera à prendre en compte.

S'il s'agit d'une "expérimentation", tous les indices biologiques seront à évaluer pour comprendre et synthétiser la connaissance contenue dans ce type de publication. Ces indices biologiques sont dispersés au sein de la publication. Le titre de l'article, l'abstract, les mots-clés de l'auteur sont des parties importantes pour comprendre l'expérimentation. En complément, des indices biologiques au sein de toutes les autres parties de cette publication ("matériel & méthodes", "résultats", "discussion", etc) pourront être utiles. Au niveau de l'introduction ou de la discussion de ces publications, on observera plutôt des citations d'autres auteurs qui nécessiteront de "retourner à l'information primaire" pour évaluer le niveau de véracité, négation, spéculation, incertitude.

S'il s'agit d'une *review*, les informations précises et intéressantes de citations de connaissances (information secondaire) d'autres auteurs dans le cadre de notre sujet, à évaluer en terme de véracité, ne seront pas toujours au niveau du titre ou du résumé mais dans une ou plusieurs zones textuelles concentrées au sein de la publication, éventuellement sous la forme de tableaux dans le corps du texte, ou en matériel supplémentaire.

Indices biologiques

Les indices biologiques (IB) relevés pourraient être classés par niveau comme par exemple "IB indispensables", "IB utiles", éventuellement "IB d'exclusion".

Il y aura comme "IB indispensable" les EN qui réfèrent aux classes qui nous intéressent : phytoplasme, vecteur, maladie, plante-hôte et les expressions qui se réfèrent à la notion de transmission ou pas.

Il pourrait y avoir comme "IB utile" les indices relatifs au cycle et au mode de vie de l'insecte vecteur ("larve", "adulte", "mâle", "femelle", "émigrant", "reémigrant", "plante refuge", "dormance", "mois", "saison", "lieu géographique", aux méthodes expérimentales (en labo, plein champ, contrôlées) aux méthodes d'identification (moléculaire, morphologique (observation)), etc.

Les "IB d'exclusion" pourraient être les EN qui ne réfèrent pas aux classes qui nous intéressent.

Au sein d'une publication pertinente, on ne trouvera pas obligatoirement la notion de "phytoplasme" car cette connaissance date de la fin des années 90. Nicolas insiste dans son HDR sur l'évolution temporelle de la connaissance des données biologiques ([Sauvion, 2022](#)). Dans le cadre de ce projet et du TAL, il me semble important d'identifier ces évolutions de nommage des EN de la catégorie "IB indispensable", de dater approximativement l'acquisition de cette connaissance. J'ai commencé à travailler sur le sujet, je souhaiterais finir ce travail avant fin juillet.

Indices linguistiques

Les informations biologiques importants sont celles qui ont traits aux phases de transmission suivantes :

- l'insecte est porteur de la bactérie pathogène ;
- l'insecte est vecteur de la bactérie pathogène ;
- l'agent pathogène induit la maladie chez la plante hôte.

Les expressions textuelles souvent relevées dans les publications à ce sujet sont par exemple :

- "une expérience de transmission"
- "virus transmission experiments"
- "a psylla-transmitted virus"
- "transmits 'a virus' capable of causing PD disease"
- "transmission of the PD agent"
- "phytoplasma passage and multiplication in the insect body comprises the latent or incubation phase"
- "infected with 'Ca. P. pyri'"

- “acquisition of ‘Ca. P. pyri”
- “detecting ‘Ca. P. pyri”
- ”infected with the PD-TW phytoplasma”
- “the pathogen that causes decline disease in pears”
- “infected by PD-TW or PD-TWII phytoplasma strains”

Diverses séquences de textes illustrant ces notions ont été mentionnées dans ce rapport (voir les études de cas ([Lemoine, 1991](#)) et sur le sujet de [Cacopsylla pyrisuga](#). J’ai aussi ébauché un tableau d’analyse de ces indices linguistiques ([voir ANNEXE n°11](#)) lors des premières semaines de mon stage. J’aurais aimé aussi avoir le temps de travailler sur un tableau de résultats d’extraction automatique d’EN pertinentes avec son contexte avant et après sur mon corpus de publication. Pour l’instant, je rencontre un problème informatique pour cette extraction. Les idées sont là, mais le temps passe trop vite.

Le cas sur le sujet de [Cacopsylla pyrisuga](#). de l’auteur [Moreno and al. \(2021\)](#), avec la phrase suivante “*Cacopsylla pyri*, *C. pyricola*, and *C. pyrisuga* (Foerster 1848; Hemiptera: Psyllidae) are the vectors of ‘Ca. P. pyri’, the causal agent PD, initially considered a virus (Jensen et al. 1964)”, illustre une citation faussée car l’information primaire de [Jensen and al. \(1964\)](#) mentionne la connaissance suivante “*C. pyrisuga* is the vectors of ‘Ca. P. pyri’, the causal agent PD, initially considered a virus”. Nous concluons cette discussion en relevant une dernière fois l’importance pour les auteurs de vérifier l’information primaire et, s’ils n’ont pas lu l’article cité, d’être plus modéré en ajoutant par exemple la notion “il semblerait”.

Conclusion personnelle

J'arrive au bout de ce rapport de stage. Il me reste encore un mois de stage sur les trois mois. Le temps est passé à "toute allure". J'ai le sentiment d'avoir établi "une bonne base de diagnostic des diverses problématiques". Il reste encore beaucoup de travail à faire quant au processus d'automatisation d'EI et des relations. L'objectif de ce rapport était pour moi de transmettre efficacement la connaissance scientifique acquise et les diverses pistes de réflexion de travail à la thésarde qui travaillera sur ce sujet à partir du mois d'octobre 2024.

J'aurais aimé fournir les résultats d'extraction automatique des contextes à partir d'Alvis NLP, aller beaucoup plus loin dans le travail sur les indices biologiques et linguistiques en particulier, avec entre autres plus d'analyse sur les modalités utilisées. J'aurais aimé aller éventuellement jusqu'à l'écriture d'un guide d'annotation sur ces indices, avoir du temps pour valoriser l'état de l'art actuel sur le sujet. J'avais commencé une bibliographie mais je n'ai pas eu le temps de l'étudier attentivement, donc je ne l'ai pas valorisée dans ce rapport. Comme je l'ai déjà mentionné dans le chapitre précédent, "les idées sont là, le temps manque".

Dès la rencontre avec C. Nédellec et N. Sauvion, le sujet m'a très grandement intéressé, y voyant un applicatif à n'importe quel domaine scientifique puisqu'il s'agit d'un travail d'analyse de publications scientifiques. Ce stage m'a aussi conforté dans l'idée que mon profil est plus celui de l'analyse de besoins et l'ingénierie des connaissances que celui de la programmation. Je tiens à rester ingénieure dans le domaine de l'information et la connaissance et non pas data-scientist ou data-analyste avec un profil très informatique.

Bibliographie

ANSES (2011). AVIS de l'Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail relatif à la réalisation d'une analyse de risque phytosanitaire sur les phytoplasmes des arbres fruitiers. *Avis de l'Anses, Saisine n° 2011-SA-013*. <https://www.anses.fr/fr/system/files/SVEG2011sa0137Ra.pdf>

Brocher F (1926). Observations biologiques sur *Psylla pyrisuga* (Hemipt.). *Ann Soc Entomol Fr.* 1926:95:183–188. <https://doi.org/10.1080/21686351.1926.12280103>

Burckhardt D., Hodkinson ID. (1986). A revision of the west Palaearctic pear psyllids (Hemiptera: Psyllidae). *Bulletin of Entomological Research.* 1986;76(1):119-132. <https://doi.org/10.1017/S0007485300015340>

Camerota C., Raddadi N., Pizzinat A. *et al.* (2012). Incidence of 'Candidatus Liberibacter europaeus' and phytoplasmas in *Cacopsylla* species (Hemiptera: Psyllidae) and their host/shelter plants. *Phytoparasitica* 40, 213–221 (2012). <https://doi.org/10.1007/s12600-012-0225-5>

Carraro L., Loi N., Ermacora P., Gregoris A. and Osler R. (1998). Transmission of Pear decline by using naturally infected *Cacopsylla pyri* L. *Acta Hort.* 472, 665-668. <https://doi.org/10.17660/ActaHortic.1998.472.89>

Carraro L., N. Loi, and P. Ermacora. (2001) The 'life cycle' of Pear decline phytoplasma in the vector "*Cacopsylla Pyri*. *Journal of Plant Pathology*, vol. 83, no. 2, 2001, pp. 87–90. JSTOR, <http://www.jstor.org/stable/41998044>

Cho G, Burckhardt D, Inoue H, Luo X, Lee S. (2017). Systematics of the east Palaearctic pear psyllids (Hemiptera: Psylloidea) with particular focus on the Japanese and Korean fauna. *Zootaxa.* 2017:4362(1):75–98. <https://doi.org/10.11646/zootaxa.4362.1.4>

Civolani S., Soroker V., Cooper W.R., Horton D.R. (2023). Diversity, biology, and management of the pear psyllids: a global look, *Annals of the Entomological Society of America*, Volume 116, Issue 6, November 2023, Pages 331–357, <https://doi.org/10.1093/aesa/saad025>

Conseil de l'Union européenne (2000). Directive 2000/29/CE du Conseil du 8 mai 2000 concernant les mesures de protection contre l'introduction dans la Communauté d'organismes nuisibles aux végétaux ou aux produits végétaux et contre leur propagation à l'intérieur de la Communauté. *Journal officiel n° L 169 du 10/07/2000* p. 0001 - 0112. <https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32000L0029>

Grbic V. 1974. Some injurious species of the family Psyllidae in pear orchards in Vojvodina. *Zastita Bilja* 25:121--31.

Hodkinson ID. (2009). Life cycle variation and adaptation in jumping plant lice (*Insecta: Hemiptera: Psylloidea*): a global synthesis. *J Nat Hist.* 2009:43(1–2):65–179. <https://doi.org/10.1080/00222930802354167>

Jarausch B., Tedeschi R., Sauvion N., Gross J., Jarausch W. (2019). Psyllid Vectors. *Phytoplasmas: Plant Pathogenic Bacteria - II : Transmission and Management of Phytoplasma - Associated Diseases*, Springer Nature, 258 p., 2019, 978-981-13-2831-2, chapitre 3. (10.1007/978-981-13-2832-9_3) https://dx.doi.org/10.1007/978-981-13-2832-9_3 . (hal-02788442) <https://hal.inrae.fr/hal-02788442>

Jensen D., Griggs W., Gonzales C., Schneider H. (1964). (a). Pear psylla proven carrier of pear decline virus. *Hilgardia* 18(3):2-3. <https://dx.doi.org/10.3733/ca.v018n03p2>

Jensen, D.D., Griggs, W.H., Gonzales, C.Q. and Schneider, H. (1964). (b). Pear decline virus transmission by pear psylla. *Phytopathology* 54: 1346-1351

Klimaszewski S. M. (1973). The jumping plant lice or psyllids (Homoptera, Psyllodea) of the Palaearctic. An annotated check-list. *Annls zool., Warsz.* 30, 155-286.

Lemoine J. (1991). Dépérissement du poirier: rôle de *Psylla pyri* dans sa dissémination. *Arboriculture fruitière* 1991:442:28–32.

Marie-Jeanne, V., Bonnot, F., Thébaud, G., Peccoud, J., Labonne, G., & Sauvion, N. (2020). Multi-scale spatial genetic structure of the vector-borne pathogen ‘*Candidatus Phytoplasma prunorum*’ in orchards and in wild habitats. *Scientific Reports*, 10(1), 5002. <https://www.nature.com/articles/s41598-020-61908-0>

Moreno A., Moreno M., Marcelo P. and Fereres A. (2021). Psyllids as major vectors of plant pathogens. *Entomologia Generalis* 5(41):419-38. <http://dx.doi.org/10.1127/entomologia/2021/1289>

République française (2014). Arrêté du 15 décembre 2014 relatif à la liste des dangers sanitaires de première et deuxième catégorie pour les espèces végétales. *Légifrance JORF n°0298 du 26 décembre 2014.* <https://www.legifrance.gouv.fr/jorf/jo/2014/12/26/0298>

Riedle-Bauer M, Paleskić C, Schönhuber C, Staples M, Brader G. (2022). Vector transmission and epidemiology of ‘*Candidatus Phytoplasma pyri*’ in Austria and identification of *Cacopsylla pyrisuga* as new pathogen vector. *J Plant Dis Prot.* 2022:129(2):375–386. <https://doi.org/10.1007/s41348-021-00526-y>

Sauvion N. (2022) Hémiptères vecteurs et prévision des risques épidémiques : à la croisée des regards. Zoologie des invertébrés. Université de Montpellier, 2022. tel-03701225.
<https://hal.science/tel-03701225v1>

Schoch CL and al. (2020) NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database (Oxford)*. [baaa062](https://doi.org/10.1093/database/baaa062). PubMed: [32761142](https://pubmed.ncbi.nlm.nih.gov/32761142/) PMC: [PMC7408187](https://pubmed.ncbi.nlm.nih.gov/PMC7408187/).

Sertkaya G., Martini M., Ermacora P. *et al.* (2005). Detection and characterization of phytoplasmas in diseased stone fruits and pear by PCR-RFLP analysis in Turkey. *Phytoparasitica* **33**, 380–390. <https://doi.org/10.1007/BF02981306>

Seljak G. (2006). An overview of the current knowledge of Jumping-Plant-Lice of Slovenia (*Hemiptera: Psylloidea*). *Acta Entomologica Slovenica vol. 14, st. 1: 11–34*.
https://www.zobodat.at/pdf/ActaEntSlov_14_0011-0034.pdf

Tedeschi R., Bosco D., Alma A. (2002). Population dynamics of *Cacopsylla melanoneura* (Homoptera: Psyllidae), a vector of apple proliferation phytoplasma in northwestern Italy. *J Econ Entomol.* 2002 Jun; 95(3):544-51. <https://doi.org/10.1146/annurev.ento.51.110104.151039>
PMID: 12075998.

Tedeschi R., Visentin C., Alma A. And Bosco D. (2003), Epidemiology of apple proliferation (AP) in northwestern Italy: evaluation of the frequency of AP-positive psyllids in naturally infected populations of *Cacopsylla melanoneura* (Homoptera: Psyllidae). *Annals of Applied Biology*, 142: 285-290. <https://doi.org/10.1111/j.1744-7348.2003.tb00252.x>

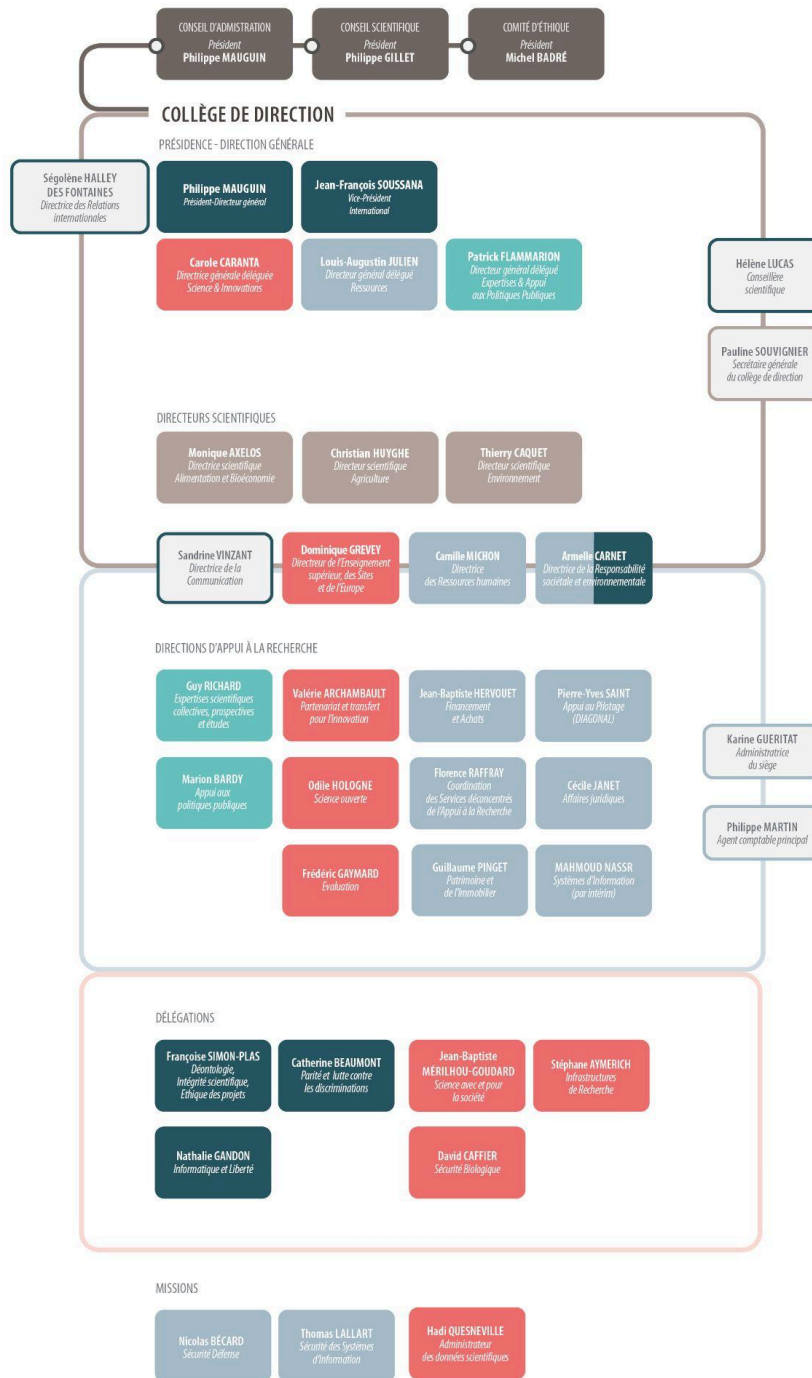
Tedeschi R., Alma A. (2004) Transmission of apple proliferation phytoplasma by *Cacopsylla melanoneura* (Homoptera: Psyllidae). *J Econ Entomol.* 2004 Feb; 97(1):8-13.
<https://doi.org/10.1111/j.1744-7348.2003.tb00252.x> PMID: 14998121.

Tedeschi R., Alma A. (2006). *Fieberiella forii* (Homoptera: Auchenorrhyncha) as a vector of 'Candidatus Phytoplasma mali'. *Plant Disease* 90, 284–290. <https://doi.org/10.1094/PD-90-0284>

Trivellone V. (2019) An online global database of Hemiptera-Phytoplasma-Plant biological interactions. *Biodiversity Data Journal* 7: e32910. <https://doi.org/10.3897/BDJ.7.e32910>

Weintraub P.G., Beanland L. (2006) Insect vectors of phytoplasmas. *Annu Rev Entomol* 51:91–111. <https://doi.org/10.1146/annurev.ento.51.110104.151039>

ANNEXE n°1 - Organigramme INRAE et présentation de l'unité de recherche MaIAGE



Source : <https://www.inrae.fr/nous-connaitre/organigramme>

Présentation de MaIAGE (Mathématiques et Informatique Appliquées du Génome à l'Environnement)

Source : <https://maiage.inrae.fr/>

L'unité de recherche (UR 1404) MaIAGE regroupe des mathématiciens, des informaticiens, des bioinformaticiens et des biologistes autour de questions de biologie et agro-écologie, allant de l'échelle moléculaire à l'échelle du paysage en passant par l'étude de l'individu, de populations ou d'écosystèmes.

Elle est structurée en cinq équipes :

- [Dynenvie](#) : modélisation dynamique et statistique pour les écosystèmes, l'épidémiologie et l'agronomie
- [Bibliome](#) : acquisition et formalisation de connaissances à partir de textes
- [BioSys](#) : biologie des systèmes
- [StatInfOmics](#) : bioinformatique et statistique des données "omiques"
- [Migale](#) : plateforme bioinformatique

L'unité développe des méthodes mathématiques et informatiques originales de portée générique ou motivées par des problèmes biologiques précis. Elle s'implique aussi dans la mise à disposition de bases de données et de logiciels permettant aux biologistes d'utiliser les outils dans de bonnes conditions ou d'exploiter automatiquement la littérature scientifique.

L'inférence statistique et la modélisation dynamique sont des compétences fortes de l'unité, auxquelles s'ajoutent la bioinformatique, l'automatique et l'algorithmique. Les activités de recherche et d'ingénierie s'appuient également sur une forte implication dans les disciplines destinataires : écologie, environnement, biologie moléculaire et biologie des systèmes.

MaIAGE est rattachée à deux départements de INRAE :

- [Mathématiques et Numérique](#) (MathNum, département pilote)
- [Microbiologie et Chaîne Alimentaire](#) (MICA)

ANNEXE n°2 - Vecteurs *Cacopsylla pyri*, *C. pyricola*, *C. pyrisuga*

'Candidatus Phytoplasma pyri' (PHYPPY)

MENU

- Overview
- Distribution
- Host plants
- Host commodities
- Vectors --
- Categorization
- Reporting
- Photos
- Documents
- Datasheet
- EPPO links

Vectors

Important note on vectors:
Information on vectors and their associated pathogens is a new feature of the database (April 2023). Data will gradually be entered by the EPPO Secretariat and will focus on regulated (quarantine) pests.

	Organism	Type
<input type="text" value="Search..."/>		- select -
	<i>Cacopsylla pyri</i> (PSYLPI)	Known vector
	<i>Cacopsylla pyricola</i> (PSYLPY)	Known vector
	<i>Cacopsylla pyrisuga</i> (PSYLPY)	Known vector

Fig n°1 : Informations sur les insectes-vecteurs de '*Candidatus Phytoplasma pyri*' en France & Europe extraites d'EPPO Global Database (<https://qd.eppo.int/taxon/PHYPPY/vectors>)

NCBI Taxonomy Browser

Search for: *Cacopsylla pyri* as complete name lock Go Clear

Display 3 levels using filter: none

Cacopsylla pyri

Taxonomy ID: 121839 (for references in articles please use NCBI:txid121839)

— current name —

Cacopsylla pyri (Linnaeus, 1758)

homotypic synonym: *Psylla pyri*

Genbank common name: **pear psyllid**

NCBI BLAST name: **psyllids**

Rank: **species**

Genetic code: [Translation table 1 \(Standard\)](#)

Mitochondrial genetic code: [Translation table 5 \(Invertebrate Mitochondrial\)](#)

Other names:

— common name(s) —

European pear sucker, pear psylla

Lineage (full)

cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Psylloidea; Psyllidae; Psyllinae; Cacopsylla

Fig n°2: Informations sur l'espèce *Cacopsylla pyri* extraites du NCBI taxonomy (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=121839>)

Cacopsylla pyri (PSYLP1)

Code created in: 2003-12-01

Overview

Basic information

- EPO Code: PSYLP1
- Preferred name: *Cacopsylla pyri*
- Authority: (Linnaeus)

Other scientific names

Name	Authority
Psylla pyri	Linnaeus

Common names

Name	Language
brunpletet pærebladloppe	Danish
gevlekte pærebladvlo	Dutch
pear psylla	English
päärynäkemppi	Finnish
Birnenblattfloh	German
Birnenblattsauger	German
braunfleckiger Birnenblattsauger	German
kriausinė blakutė	Lithuanian
liten pæresuger	Norwegian
psila-da-pereira	Portuguese
psila de la pera	Spanish
psilido del peral	Spanish
mindre päronbladlöppla	Swedish

Taxonomy

- Kingdom: Animalia (1ANIMK)
- Phylum: Arthropoda (1ARTHPO)
- Subphylum: Hexapoda (1HEXAQ)
- Class: Insecta (1INSEC)
- Order: Hemiptera (1HEMIO)
- Suborder: Sternorrhyncha (1STERR)
- Family: Psyllidae (1PSYLF)
- Genus: *Cacopsylla* (1CCPSG)
- Species: *Cacopsylla pyri* (PSYLP1)

more photos...

Fig n°3 : Informations sur l'espèce *Cacopsylla pyri* extraites d'EPPO Global Database (<https://gd.eppo.int/taxon/PSYLP1>)

Classification

- Kingdom: Animalia
- Phylum: Arthropoda
- Class: Insecta
- Order: Hemiptera
- Family: Psyllidae
- Genus: *Cacopsylla* Osborn, 1910
- Species: *Cacopsylla pyri* (Linnaeus, 1758)
 - *Cacopsylla pyri* (Linnaeus, 1758)
 - *Psylla pyri* (Linnaeus, 1758)
 - *Psylla pyri* (Linnaeus, 1758)

Species

Cacopsylla pyri (Linnaeus, 1758)

FAUNE EUROPEA

56 OCCURRENCES AVEC DES IMAGES

27 OBSERVATIONS GÉOLOCALISÉES

Fig n°4 : Extrait de Fauna Europaea <https://www.gbif.org/fr/species/2012888>

NCBI Taxonomy Browser

Entrez PubMed Nucleotide Protein Genome Structure PMC

Search for *Cacopsylla pyricola* as complete name lock Go Clear

Display 3 levels using filter: none

Cacopsylla pyricola

Taxonomy ID: 191317 (for references in articles please use NCBI:txid191317)

current name

Cacopsylla pyricola (Forster, 1848)

NCBI BLAST name: **psyllids**

Rank: **species**

Genetic code: [Translation table 1 \(Standard\)](#)

Mitochondrial genetic code: [Translation table 5 \(Invertebrate Mitochondrial\)](#)

[Lineage \(full\)](#)

cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Psylloidea; Psyllidae; Psyllinae; *Cacopsylla*

Fig n°5 : Informations sur l'espèce *Cacopsylla pyricola* extraites du NCBI taxonomy (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=191317>)

Cacopsylla pyricola (PSYLPC)

Code created in: 2000-12-25

MENU

- Overview --
- Distribution
- Vector of
- Categorization
- Photos
- Documents

Overview

Basic information

- EPPO Code: PSYLPC
- Preferred name: *Cacopsylla pyricola*
- Authority: (Forster)

Other scientific names


Name	Authority
<i>Psylla piricola</i>	
<i>Psylla pyricola</i>	Forster
<i>Psylla simulans</i>	

Common names

Name	Language
gul pærebladloppe	Danish
perebladvlo	Dutch
perenbladvlo	Dutch
pear psylla	English
pear psyllid	English
pear sucker	English
pieni päärynäkemppi	Finnish
psylle du poirier	French
Birnenblattfloh	German
Birnenblattsauger	German
gelber Birnenblattsauger	German
psilla del pero	Italian
gul pæresuger	Norwegian
pulgullia del peral	Spanish
psillido del peral	Spanish (AR)
lilla päronbladloppan	Swedish
armut yaprak piresi	Turkish

Taxonomy

- Kingdom: Animalia (1ANIMK)
- Phylum: Arthropoda (1ARTHYP)
- Subphylum: Hexapoda (1HEXAQ)
- Class: Insecta (1INSEC)
- Order: Hemiptera (1HEMIO)
- Suborder: Sternorrhyncha (1STERR)
- Family: Psyllidae (1PSYLF)
- Genus: *Cacopsylla* (1CCPSG)
- Species: *Cacopsylla pyricola* (PSYLPC)



more photos...

Fig n°6 : Informations sur l'espèce *Cacopsylla pyricola* extraites d'EPPO Global Database (<https://gd.eppo.int/taxon/PSYLPC>)

Search for: as complete name lock

Display: levels using filter:

Cacopsylla pyrisuga

Taxonomy ID: 1191701 (for references in articles please use NCBI:txid1191701)

current name
Cacopsylla pyrisuga (Foerster, 1848)

NCBI BLAST name: **psyllids**
 Rank: **species**
 Genetic code: [Translation table 1 \(Standard\)](#)
 Mitochondrial genetic code: [Translation table 5 \(Invertebrate Mitochondrial\)](#)

Lineage(full)
[cellular organisms](#); [Eukaryota](#); [Opisthokonta](#); [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Protostomia](#); [Ecdysozoa](#); [Panarthropoda](#); [Arthropoda](#); [Mandibulata](#); [Pancrustacea](#); [Hexapoda](#); [Insecta](#); [Dicondylia](#); [Pterygota](#); [Neoptera](#); [Paraneoptera](#); [Hemiptera](#); [Sternorrhyncha](#); [Psylloidea](#); [Psyllidae](#); [Psyllinae](#); [Cacopsylla](#)

Fig n°7 : Informations sur l'espèce *Cacopsylla pyrisuga* extraites du NCBI taxonomy
 (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=1191701>)

Overview

Basic information

- EPPO Code: PSYLPY
- Preferred name: *Cacopsylla pyrisuga*
- Authority: (Förster)

Other scientific names

Name	Authority
<i>Psylla austriaca</i>	Flor
<i>Psylla pyrisuga</i>	Förster
<i>Psylla rufitarsis</i>	Meyer-Dür
<i>Psylla rutila</i>	Meyer-Dür

Common names

Name	Language
stor pærebladloppe	Danish
grote perebladvlo	Dutch
pear sucker	English
iso päärynäkemppi	Finnish
psylle du poirier	French
großer Birnenblattsauger	German
nasi-kizirami	Japanese
stor pæresuger	Norwegian
psilla del peral	Spanish
större päronbladloppa	Swedish

Taxonomy

- Kingdom: Animalia (1ANIMK)
- Phylum: Arthropoda (1ARTHYP)
- Subphylum: Hexapoda (1HEXAQ)
- Class: Insecta (1INSEC)
- Order: Hemiptera (1HEMIO)
- Suborder: Sternorrhyncha (1STERR)
- Family: Psyllidae (1PSYLF)
- Genus: *Cacopsylla* (1CCPSG)
- Species: *Cacopsylla pyrisuga* (PSYLPY)

Fig n°8 : Informations sur l'espèce *Cacopsylla pyrisuga* extraites d'EPPO Global Database
 (<https://gd.eppo.int/taxon/PSYLPY>)

ANNEXE n°3 - Script Alvis NLP “main.plan” que j’ai fréquemment utilisé

```
<alvisnlp-plan id="main">

  <read class="TikaReader">
    <source>documents</source>
  </read>

  <ne>
    <taxo>
      <detect href="taxa.plan">
        <taxaDict>resources/taxo/taxa+id_full.txt</taxaDict>
        <compiledDict>resources/taxo/taxa+id_full.trie</compiledDict>
      </detect>

      <overlaps class="RemoveOverlaps">
        <layer>taxa</layer>
        <removeEqual>false</removeEqual>
      </overlaps>

      <Plante href="select-taxa.plan">
        <name>Plante</name>
        <taxidFile>resources/taxo/taxid_seed-plants.txt</taxidFile>
      </Plante>

      <Vect href="select-taxa.plan">
        <name>Vect</name>
        <taxidFile>resources/taxo/taxid_vecteurs.txt</taxidFile>
      </Vect>

      <ON href="select-taxa.plan">
        <name>ON</name>
        <taxidFile>resources/taxo/taxid_ON.txt</taxidFile>
      </ON>
    </taxo>

    <stanford class="StanfordNER">
      <classifierFile>resources/stanford/english.all.3class.distsim.crf.ser.gz</classifierFile>
      <labelFeature>stanford-label</labelFeature>
      <targetLayer>stanford</targetLayer>
    </stanford>
  </ne>
</alvisnlp-plan>
```

```

<searchInContents/>
</stanford>

<geo-layer class="Action">
  <target>documents.sections.layer:stanford[@stanford-label == "LOCATION" and not
span:taxa]</target>
  <action>add:Geographical|set:feat:type("Geographical")</action>
  <addToLayer/>
  <setFeatures/>
</geo-layer>

<org-layer class="Action">
  <target>documents.sections.layer:stanford[@stanford-label == "ORGANIZATION" and
not span:taxa]</target>
  <action>add:Organization|set:feat:type("Organization")</action>
  <addToLayer/>
  <setFeatures/>
</org-layer>

<person-layer class="Action">
  <target>documents.sections.layer:stanford[@stanford-label == "PERSON" and not
span:taxa]</target>
  <action>add:Person|set:feat:type("Person")</action>
  <addToLayer/>
  <setFeatures/>
</person-layer>

<rigid-entities class="MergeLayers">
  <sourceLayers>taxa,Geographical,Organization,Person</sourceLayers>
  <targetLayer>rigid-entities</targetLayer>
</rigid-entities>
</ne>

<seg href="res://segmentation.plan"/>

<tt class="TreeTagger">
  <noUnknownLemma/>
</tt>

<http class="QuickHTML">
  <outDir>output/html_nouveau</outDir>

<mentionLayers>ON,Plant,Vect,Disease,Geographical,Organization,Person,Citation</mentionL
ayers>

```

```

<typeFeature>type</typeFeature>
<layoutLayer>html</layoutLayer>
<features>type,taxid,canonical-name</features>
<colorMap>
<ON>#A4C2F4</ON>
<Vect>#D5A6BD</Vect>
<Plant>#B6D7A8</Plant>
<Disease>#EA9999</Disease>
<Geographical>#B7B7B7</Geographical>
<Organization>#919090</Organization>
<Person>#F6B379</Person>
<Citation>#B48CBA</Citation>
</colorMap>
</http>

<table class="TabularExport">
  <outDir>./</outDir>
  <corpusFile>output/entities.csv</corpusFile>

<lines>documents.sections.(layer:ON|layer:Plant|layer:Vect|layer:Disease|layer:Geographical|la
yer:Organization|layer:Person|layer:Citation)</lines>
  <columns separator=";">
    section.document.@id;
    @type;
    int(start) ^ "-" ^ int(end);
    @form;
    @taxid;
    @canonical-name
  </columns>
</table>
</alvisnlp-plan>

```

Partie rajoutée dans un deuxième temps :

```

<table class="TabularExport">
  <outDir>./</outDir>
  <corpusFile>output/entities.csv</corpusFile>

<lines>documents.sections.(layer:ON|layer:Plant|layer:Vect|layer:Disease|layer:Geographical|la
yer:Organization|layer:Person|layer:Citation)</lines>
  <columns separator=";">
    section.document.@id;

```

```
@type;  
int(start) ^ "-" ^ int(end);  
ctx:words-before(10); <!-- ctx:sentence-before -->  
@form;  
ctx:words-after(10); <!-- ctx:sentence-after -->  
@taxid;  
@canonical-name  
</columns>  
</table>
```

ANNEXE n°4 - Dictionnaire du phytoplasme considéré et des divers nommages

Candidatus Phytoplasma pyri

Ca. P. pyri

'Candidatus Phytoplasma pyri'

Phytoplasma pyri

bacteria

phytoplasma

phytopathogenic agents

pathogen

pathogen transmission

phytoplasma 'Candidatus Phytoplasma pyri'

phytoplasmas responsible for diseases

the phytoplasma 'Candidatus Phytoplasma pyri' (CaPpyri)

CaPpyri

phytoplasma 'Candidatus P.pyri'

phytoplasma CaPpyri

Pear decline VIRUS

virus

a psylla-transmitted virus

Pear decline phytoplasma

Pear decline phy-toplasma

causal agent

phytoplasma belonging to the apple proliferation group

'a virus' capable of causing PD disease

transmission of the PD agent

PD phyto-plasma

PD phytoplasma

presence of PD phytoplasma

'Candidatus Phytoplasma' species

one group of phytoplasmas, the 16SrX or apple proliferation group

16SrX

apple proliferation group

phytoplasmas

phytoplasma-infected trees

phytoplasmas of the group 16SrX

phloem-limited

phytoplasma passage and multiplication in the insect body comprises the latent or incubation phase

phloem-limited

phytoplasmas are phloem-limited

virus-borne

'Ca. P. pyri'

acquisition of 'Ca. P. pyri'

number of phytoplasmas per psyllid

'Ca. P. pyri'

detecting 'Ca. P. pyri'

detect 'Ca. P. pyri'

'Ca. P. pyri' transmission by

'Ca. P. pyri'

infected with the PD-TW phytoplasma

the pathogen that causes decline dis-ease in pears

infected by PD-TW or PD-TWII phytoplasma strains

PYLR phytoplasma

Experimental transmission of PYLR phyto-plasma to

“Ca. Phytoplasma pyri”

Known vectors of Phytoplasma pyri

The pathogen is found in pear-growing regions

Phytoplasma pyri

phytoplasma-infected C. pyricola

Ca. Phytoplasma pyri

Ca.P.pyri-PD-AJ542543

Candidatus Phytoplasma pyri

PDTW phytoplasma

transmit phytoplasmas

Pear decline phytoplasmas

16SrX-C

Pear decline virus transmission by

‘Candidatus Phytoplasma pyri’ strain PD-TW

strain PD-TW

PDTW phytoplasma

PYLR phytoplasma

ANNEXE n°5 - Dictionnaire des psylles considérés et des divers nommages

psyllids as vectors of bacteria in plant

pear psyllids

PEAR PSYLLA

transmitted by the pear psylla

psyllids vectors of a phytoplasma on pear

insect of the piercing-sucking type

the pear psylla (*Psyllu pyricolu* Foerster - correction : *Psylla pyricola* Foerster)

pear psylla ;

Hemiptera

sap-sucking piercing-sucking insects

Psyllidae

adult psylla

Cacopsylla Ossiannilsson

psyllids

phloem feeding hemipterans

Psylloidea

planthoppers

Fulgoroidea

Auchenorrhyncha

Sternorrhyncha

psylla-transmitted virus

vector *Cacopsylla pyri*

overwintering adults of *C. pyri*

Psylla pyri-cola Förster

Psylla pyricola Förster

Cacopsylla pyricola Förster

pear psylla (*Psylla pyri-cola* Förster, now *Cacopsylla pyricola* Förster)

Psylla pyri-cola

vector *C. pyricola*

in *C. pyri*

Psyllid Vectors

genus *Cacopsylla*

vector species

Psyllid vectors

Cacopsylla species

psyllids

phloem feeders

vector of PYLR phytoplasma

phloem-feeding insects

feeding in the phloem of infected plants

Pear Psyllids

recognized vectors of Pear decline

presumed vectors of Pear decline

the transmission capacity of *C. chinensis*

exposed to *C. chinensis* specimens

Insect vectors for

psyllid species

pear psyllids (*Cacopsylla* Ossiannilsson; Hemiptera: Psylloidea: Psyllidae)

pear psyllids

The pear psyllids (Hemiptera: Psylloidea: *Cacopsylla*)

Pear psyllids

species of pear psyllid

Cacopsylla pyricola (Foerster)

Cacopsylla pyri (L.).

The pear psyllids (*Cacopsylla* spp.; Psyllidae)

Cacopsylla species

pear psyllids

vectored by adult pear psyllids

European pear psyllids

Asian pear psyllid

pear psyllids

Pear suckers

Hemiptera

phloem-feeding insects

membracids

Cacopsylla spp.

Membracidae

Ca. pyri (Linnaeus)

Ca. pyricola (Forster)

Cacopsylla pyrisuga (Forster)

Cacopsylla pyri L.

pear psylla

Psylla pyri

Cacopsylla nigella-group

Cacopsylla pyri-group

Cacopyrisuga-group

Psylla Pyrisuga

Psylla pyri

P. pyri

Pysylla pyri

Psylla pyricola

P. pricola

Psylla pyricolla

Psylla piri

Psylla piri L.

P. pyri L.

Psylla piricola

Psylla piricola Foerst.

P. pyricola Foerst.

Psylla pirisuga Foerst.

Psylla pirisuga

P. pirisuga Foerst.

Psyllids

Cacopsylla (Hepatopsylla) notata (Flor)

Cacopsylla (Thamnopsylla) pyrisuga (Förster)

Cacopsylla (Hepatopsylla) pyri (L.)

Cacopsylla pyri

Cacopsylla pyri-group

PSYLPI

Cacopsylla pyri-group

C. pyri

Cacopsylla pyri-group

Cacopsylla pyricola

Cacopsylla pyri-group

PSY LPC

Cacopsylla pyri-group

C. pyricola

Cacopsylla pyri-group

Cacopsylla bidens

Cacopsylla pyri-group

C. bidens

Cacopsylla pyri-group

Psylla vasilevi

Cacopsylla pyri-group

Psylla jiangli

Cacopsylla pyri-group

Cacopsylla notata

Cacopsylla pyri-group

C. notata

Cacopsylla pyri-group

Cacopsylla betulaefoliae

Cacopsylla pyri-group

C. betulaefoliae

Cacopsylla pyri-group

Psylla heterobetulaefoliae

Cacopsylla pyri-group

Cacopsylla chinensis

Cacopsylla pyri-group

<i>C. chinensis</i>	<i>Cacopsylla pyri</i> -group
<i>Cacopsylla guangdongli</i>	<i>Cacopsylla pyri</i> -group
<i>C. guangdongli</i>	<i>Cacopsylla pyri</i> -group
<i>Cacopsylla donggangica</i>	<i>Cacopsylla pyri</i> -group
<i>C. donggangica</i>	<i>Cacopsylla pyri</i> -group
<i>Cacopsylla fera</i>	<i>Cacopsylla pyri</i> -group
<i>C. fera</i>	<i>Cacopsylla pyri</i> -group
<i>Cacopsylla flavianthracina</i>	<i>Cacopsylla pyri</i> -group
<i>C. flavianthracina</i>	<i>Cacopsylla pyri</i> -group
<i>Cacopsylla gyrogenna</i>	<i>Cacopsylla pyri</i> -group
<i>C. gyrogenna</i>	<i>Cacopsylla pyri</i> -group
<i>Cacopsylla jukyungi</i>	<i>Cacopsylla pyri</i> -group
<i>C. jukyungi</i>	<i>Cacopsylla pyri</i> -group
<i>Cacopsylla maculatili</i>	<i>Cacopsylla pyri</i> -group
<i>C. maculatili</i>	<i>Cacopsylla pyri</i> -group
<i>Cacopsylla qiuzili</i>	<i>Cacopsylla pyri</i> -group
<i>C. qiuzili</i>	<i>Cacopsylla pyri</i> -group
<i>Cacopsylla nervinigra</i>	<i>Cacopsylla pyri</i> -group
<i>C. nervinigra</i>	<i>Cacopsylla pyri</i> -group
<i>Cacopsylla permixta</i>	<i>Cacopsylla pyri</i> -group
<i>C. permixta</i>	<i>Cacopsylla pyri</i> -group

Cacopsylla Pyri (Linné, 1758)	Cacopsylla pyri-group
Cacopsylla Pyri (Linnaeus, 1758)	Cacopsylla pyri-group
Apiopsylla	Cacopsylla pyri-group
Psylla pyrarboris	Cacopsylla pyri-group
Cacopsylla Pyricola (Foerster, 1848)	Cacopsylla pyri-group
Psylla apiophila	Cacopsylla pyri-group
Psylla argyrostigma	Cacopsylla pyri-group
Psylla simulans	Cacopsylla pyri-group
Psylla horvathii	Cacopsylla pyri-group
Psylla vicina	Cacopsylla pyri-group
Cacopsylla xanthisma	Cacopsylla pyri-group
C. xanthisma	Cacopsylla pyri-group
Cacopsylla yunli	Cacopsylla pyri-group
C. yunli	Cacopsylla pyri-group
Cacopsylla Pyricola	Cacopsylla pyri-group
betulaefoliae	Cacopsylla pyri-group
het-erobetulaefoliae	Cacopsylla pyri-group
qiuzili	Cacopsylla pyri-group
maculatili	Cacopsylla pyri-group
mac-ulatili	Cacopsylla pyri-group
cinereosignata	Cacopsylla pyri-group

jukyungi	Cacopsylla pyri-group
Cacopsylla pyrisuga	Cacopsylla pyrisuga-group
PSYLPY	Cacopsylla pyrisuga-group
C. pyrisuga	Cacopsylla pyrisuga-group
Psylla austriaca	Cacopsylla pyrisuga-group
Psylla aurantiaca	Cacopsylla pyrisuga-group
Psylla rufitarsis	Cacopsylla pyrisuga-group
Psylla rutila	Cacopsylla pyrisuga-group
Cacopsylla burckhardti	Cacopsylla pyrisuga-group
C. burckhardti	Cacopsylla pyrisuga-group
Cacopsylla melanoneura	Cacopsylla pyrisuga-group
C. melanoneura	Cacopsylla pyrisuga-group
Cacopsylla accincta	Cacopsylla pyrisuga-group
C. accincta	Cacopsylla pyrisuga-group
burckhardti	Cacopsylla pyrisuga-group

ANNEXE n°6 - Dictionnaire sur la maladie et ses divers nommages

Pear decline

This disease, known as Pear decline (PD)

decline disease of pears

quick decline symptoms (wilting and collapse of foliage and succulent shoots) developed in several of the trees ;

causes the sudden wilting and collapse of pear trees

Pear decline disease

dangerous dis-eases of pear trees

The disease has long been known as 'moria' in northern Italy (Refatti, 1948)

PD

dis-eases of pear trees

causing PD disease

vector's host plant

fruit crop diseases associated with the presence of phytoplasmas

diseases of fruit crops

diseases of pear trees

PD-TW

PD-TWII

decline dis-ease in pears, shown by yellowing of leaves, decreased tree vigor, and death of the tree

the causal agents of Pear decline disease

pear trees with decline symptoms

Deperissement du poirier

ANNEXE n°7 - Dictionnaire sur la plante hôte et ses divers noms

pear

pear psyllids

their host plant

pear orchards

pear trees (Pyrus)

PEAR

pears

orchard plots

Pyrus serotina

pear trees

PD-affected orchards

infect test plants

Pyrus communis L.

their host plants

Rosaceae

fruit trees

fruit crops

their host plants

live on pear

P. communis

Pyrus species

conifers and remigrate to *Pyrus*

Pyrus pyrifolia

P. pyrifolia

Pyrus ussuriensis

P. ussuriensis

Pyrus amygdaliformis

P. amygdaliformis

Pyrus pashia

P. pashia

Pyrus salicifolia

P. salicifolia

Pyrus betulifolia

P. betulifolia

Pyrus pyraaster

P. pyraaster

Pyrus syriaca

P. syriaca

Pyrus bretschneideri

P. bretschneideri

Pyrus pyrifolia

P. pyrifolia

Pyrus sp.

Pyrus bucharica

P. bucharica

Pyrus eleagrifolia

P. eleagrifolia

Pyrus calleryana

P. calleryana

Pyrus phaeocarpa

P. phaeocarpa

from experimentally infected pear seedlings

the upper part of the tree from the roots

dormant plants

aerial plant parts

from infected pears to healthy periwinkles

peach seedlings

Pyrus

host to pear

hosts to pear

hosts to pear (Rosaceae: *Pyrus*)

intensive pear cultivation

pear orchards

pear producers

producers of pears

in pears

European pear

P. betulifolia Bunge

P. calleryana Decne.

P. ussuriensis Maxim.

pear orchards ;

Pyrus spp.

serious pests on cultivated pear

cultivated pear

poirier

ANNEXE n°8 - Dictionnaire de verbes et marqueurs cités et intéressants dans le cadre de cette étude

to be able

ability of vector

acquire

affect

aimed

are vector

are vectored by

associated

became infected

been found

behavior

being associated

burns

but the relationship between this vector and the PD phytoplasma has yet to be clarified

can

can not

can potentially acquire and transmit

cannot

causal

cause

cause direct

causer
causing
changes
co-infected
competency
concluded
confirmed
confused
confusing
confusion
consider valid
could
could happen
could not transmit
damage
decline
disseminated
downgrading
elucidate
enable
enabled to proof
experiments
found
happen
have probably been

improve
improved
improving
infected
ingest
inoculated
is caused by
is confirmed
limited
located
losses
may
may cause
may improve
may taste
misidentified
not being associated with
not transmit
not yet confirmed
penetrate
potentially
presumably
previously misidentified
proof
prove

proved

proved to be

provide

provide very strong evidence

recognized vectors

reported

responsible for diseases

retained

reviewed

secreted

show

spread

still needs to be proven

sucking

transmission

transmit

transmit phytoplasmas

transmitted

untangling

untangling the confusing

vector transmission

vector competency

vectored

was aimed

was found

would

transmission experiments

a psylla-transmitted virus

transmits 'a virus' capable of causing PD disease

transmission of the PD agent

phytoplasma passage and multiplication in the insect body comprises the latent or incubation phase

infected with 'Ca. P. pyri'

acquisition of 'Ca. P. pyri'

detecting 'Ca. P. pyri'

infected with the PD-TW phytoplasma

the pathogen that causes decline disease in pears

infected by PD-TW or PD-TWII phytoplasma strains

vector transmission

identification ... as new pathogen vector

the vectoring ability

to be experimentally proven for the first time

can transmit

ANNEXE n°9 - Script conçu sur la base des dictionnaires personnalisés

```
<?xml version="1.0" encoding="UTF-8"?>

<alvisnlp-plan id="project-on-txt">
  <read class="TikaReader">
    <source>documents</source>
  </read>

  <ne>
    <project1 class="TabularProjector">
      <dictFile>dic_psyllle.txt</dictFile>
      <targetLayer>nom</targetLayer>
      <valueFeatures>name,groupe</valueFeatures>
      <constantAnnotationFeatures>type=Dico_psyllle</constantAnnotationFeatures>
    </project1>

    <project2 class="TabularProjector">
      <dictFile>dic_phytoplasme.txt</dictFile>
      <targetLayer>nom</targetLayer>
      <valueFeatures>name,groupe</valueFeatures>
      <constantAnnotationFeatures>type=Dico_phytoplasme</constantAnnotationFeatures>
    </project2>

    <project3 class="TabularProjector">
      <dictFile>dic_plant_hote.txt</dictFile>
      <targetLayer>nom</targetLayer>
      <valueFeatures>name,groupe</valueFeatures>
      <constantAnnotationFeatures>type=Dico_plante_hote</constantAnnotationFeatures>
    </project3>

    <project4 class="TabularProjector">
      <dictFile>dic_maladie.txt</dictFile>
      <targetLayer>nom</targetLayer>
      <valueFeatures>name,groupe</valueFeatures>
      <constantAnnotationFeatures>type=Dico_maladie</constantAnnotationFeatures>
    </project4>

    <project5 class="TabularProjector">
```

```

    <dictFile>dic_pays.txt</dictFile>
    <targetLayer>nom</targetLayer>
    <valueFeatures>name,groupe</valueFeatures>
    <constantAnnotationFeatures>type=Dico_pays</constantAnnotationFeatures>
</project5>

<project6 class="TabularProjector">
    <dictFile>dic_verbe.txt</dictFile>
    <targetLayer>nom</targetLayer>
    <valueFeatures>name,groupe</valueFeatures>
    <constantAnnotationFeatures>type=Dico_verbe</constantAnnotationFeatures>
</project6>

<stanford class="StanfordNER">
    <classifierFile>resources/stanford/english.all.3class.distsim.crf.ser.gz</classifierFile>
    <labelFeature>stanford-label</labelFeature>
    <targetLayer>stanford</targetLayer>
    <searchInContents/>
</stanford>

<person-layer class="Action">
    <target>documents.sections.layer:stanford[@stanford-label == "PERSON" and not
span:taxa]</target>
    <action>add:Person|set:feat:type("Person")</action>
    <addToLayer/>
    <setFeatures/>
</person-layer>
</ne>

<seg href="res://segmentation.plan"/>

<tt class="TreeTagger">
    <noUnknownLemma/>
</tt>

<http class="QuickHTML">
    <outDir>output/html_projection</outDir>
    <mentionLayers>nom,Geographical</mentionLayers>
    <typeFeature>type</typeFeature>
    <layoutLayer>html</layoutLayer>
    <features>type</features>
    <colorMap>
        <Dico_psylle>#D5A6BD</Dico_psylle>

```

<Dico_phytoplasme>#A4C2F4</Dico_phytoplasme>
<Dico_plante_hote>#B6D7A8</Dico_plante_hote>
<Dico_maladie>#EA9999</Dico_maladie>
<Dico_pays>#B7B7B7</Dico_pays>
<Dico_verbe>#F6B379</Dico_verbe>
<Person>#EE82EE</Person>
</colorMap>
</http>
</alvisnlp-plan>

ANNEXE n°10 - Analyse détaillée concernant l'espèce "*Cacopsylla pyrisuga*"

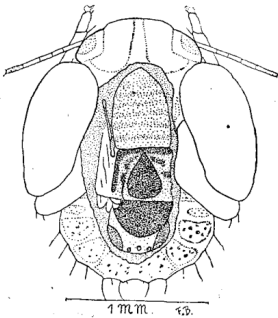
1926 : Brochet

Observation terrain, description ; 3 espèces sur le poirier, dépend de la saison ; les nymphes infestées (changement de couleur) de *P. Pyrisuga* disparaissent mystérieusement

Voilà le schéma de l'existence visible et connue des *Psylla pyrisuga*.

Cette existence commence par un mystère : d'où viennent ces insectes — que l'on n'observe que pendant quelques semaines (10 à 12, au maximum)? elle se termine par un autre mystère : que deviennent ils, à peine apparus?

Or, pendant ces quelques semaines, leur vie est riche en incidents et en ... accidents.



Psylla pyrisuga. — Nymph, avec l'Hyménoptère parasite.

Les phénomènes que je viens de relater se rapportent à *Psylla pyrisuga*, qu'on trouve, au printemps, sur les poiriers.

En été et en automne, on trouve d'autres Psylles sur ces mêmes arbres; ce sont *Psylla pyri* et *P. pricolla*. Ils sont de plus petite taille et leurs larves ne forment pas de colonies; on les trouve fixées à l'aisselle des feuilles, à l'état d'individus isolés.

1973 : Klimaszewski

Taxonomie / Distribution : noter le commentaire "Whenever appears in larger quantity *P. pyrisuga* may be an important pest in orchards"

329. *Psylla pyrisuga* FÖRSTER, 1848.*Psylla pyrisuga*: AULMANN 1913.

Distribution. Whole Palaearctic (DOBREANU, MANOLACHE 1962, HARISANOV, LAUTERER 1968, HAUPT 1935a, HORVÁTH 1918, JACOBSEN 1919, JOAKIMOV 1909, KLIMASZEWSKI 1965b, 1967a, KLOET, HINCKS 1964, LOGINOVA 1964a, 1966b, MIYATAKE 1964a, 1969, NOVAK, WAGNER 1962, OSSIANNILSSON 1952, OSSIANNILSSON, LINDBERG 1960, SASAKI 1954, SCHAEFER 1949, VONDRAČEK 1953a, 1957a, WAGNER, FRANZ 1961).

Host plant. *Pirus communis*.Whenever appears in larger quantity *P. pyrisuga* may be an important pest in orchards.

Renvoi vers cette référence bibliographique :

BROCHER F. 1926. Observations biologiques sur *Psylla pyrisuga*. Ann. Soc. ent. France, Paris, **95**: 183 – 188.

1974 : Grbic (document compliqué à obtenir + langue croate + mauvaise qualité de photocopie)Titre traduit en fin d'article : "Some nuisible species of *Psylla* spp., on the pear plantations in Vojvodina"

Résumé en anglais disponible à la fin de l'article :

"A long-term study has revealed that **the pear crop in the area is parasided by two pest species from the genus *Psylla*, viz *P. piri* and *P. pirisuga* Foerst.** In pear orchards of Vojvodina ***P. piri* is the dominant species, while *P. pirisuga* occurs only sporadically,** always in association with the former species, and confined to certain localities.

The research has shown that under the given environmental conditions ***P.piri* has five full generations a year, while *P. pirisuga* develops only one generation on pear and then moves to other hosts.**

Apart from the immediate damage which these lice can cause, **they are very important as vectors of the disease known as Pear decline,** which we first registered in Yugoslavia in 1966."

1986 : Burckhardt & HodkinsonC. *pyrisuga* (Förster).

(Abstract) "taxonomy ... Pyrus-feeding ... Genus *Cacopsylla* ... **7 espèces** dont *C. pyrisuga* (Forster)"

Analyse morphologique : descriptions and identification keys

"... It is easily separated from *C. pyrisuga* by its smaller size.

Renvoi entre autres à la référence de **Klimaszewski, 1973** :

"**Eight species of Pyrus-feeding psyllids**, here referred to the genus *Cacopsylla* Ossiannilsson, were recognized from the Palaearctic Region west of China prior to this present study (Klimaszewski, 1973; Sulc, 1915). Several other species had been sunk in synonymy (Hodkinson, 1984). The recognized species were *C. pyricola* (Forster), *C. pyri* (L.), ***C. pyrisuga* (Forster)**, *C. vasilevi* (Sulc), *C. vicina* (Sulc), *C. bidens* (Sulc), *C. horvathii* (Sulc) and *C. fera* (Baeva). *Psylla picta* Forster also appears to belong to this group. *C. pyricola* is recorded as an introduction into both North and South America and, together with ***C. pyrisuga*** and *C. pyri*, is recorded from the temperate far east. The published geographical distributions and host-plant records of all the above species are summarized by Hodkinson (1984)."

1991 : Armand

L'hyperparasitisme, essentiellement dû à *Syrphophagus mamitus* (Walker), apparaît plus tard sur l'hôte-relais *Psylla pyrisuga* Foerster présent en vergers entre la 1^{re} et la 2^e génération de *P. pyri*.

Le « psylle commun du poirier » *Psylla pyri* (L.) est le ravageur-clé de cette culture en France et dans de nombreux pays d'Europe. Parmi les antagonistes, nous avons étudié l'influence du complexe parasitaire sur la dynamique des populations du ravageur.

Ce travail a recours au contrôle visuel des rameaux pour suivre le niveau des populations du phytophage. Une méthode d'isolement de larves de l'hôte au laboratoire est utilisée afin d'évaluer le taux de parasitisme ou prévalence.

La parasitofaune est constituée de micro-hyménoptères parasitoïdes et hyperparasitoïdes de larves L4 et L5 de psylle. *Trechmites psyllae* (Ruschka) et dans une moindre mesure *Prionomitus mitratus* (Dalman) sont les parasitoïdes qui assurent la régulation dès la 1^{re} génération du ravageur. L'hyperparasitisme, essentiellement dû à *Syrphophagus mamitus* (Walker), apparaît plus tard sur l'hôte-relais *Psylla pyrisuga* Foerster présent en vergers entre la 1^{re} et la 2^e génération de *P. pyri*. Les autres hyperparasitoïdes, *Pachyneuron muscarum* (L.), *Pachyneuron aphidis* Bouche, *Dendrocerus psyllarum* Dessart, *Dilyta subclavata* Foerster et *Dilyta talitzkii* Belizin sont encore plus tardifs.

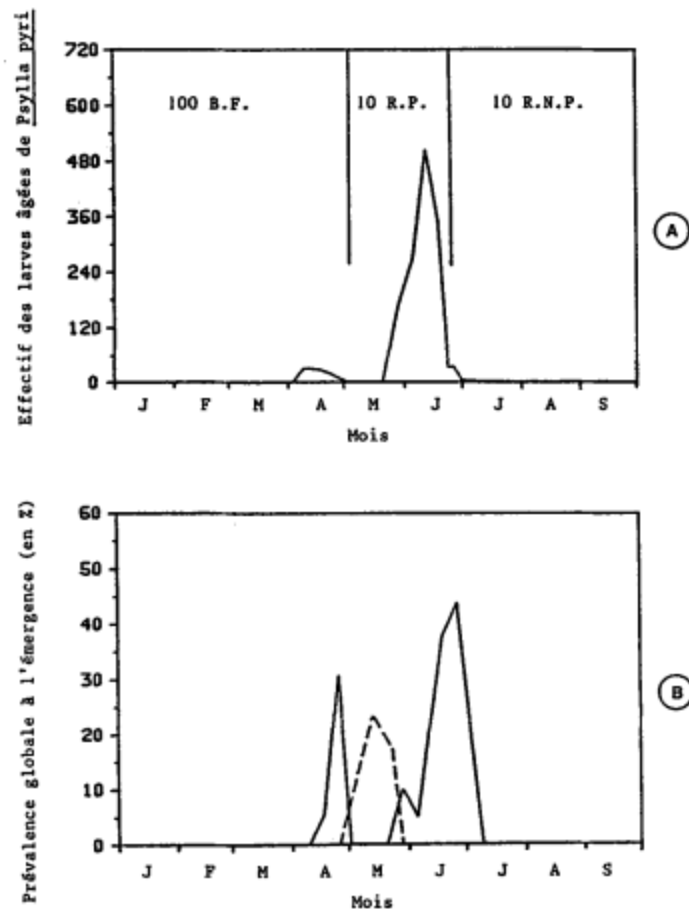


Fig. 1A. Evolution, dans le verger Poirson en 1985, de l'effectif des larves âgées (L4, L5) de *Psylla pyri* dénombrées par le contrôle visuel de 100 lambourdes évoluant en bouquets floraux puis fruitiers (100 B.F.) jusqu'au 02.05 puis de 10 rameaux poussants (10 R.P.) jusqu'au 24.06 et ensuite de 10 rameaux non poussants (10 R.N.P.).

Fig. 1B. Evolution de la prévalence globale à l'émergence de *Psylla pyri* (—) et de *Psylla pyrisuga* (----) suivie par encagement de 100 larves âgées de l'hôte par date de prélèvement dans le verger Poirson en 1985.

2006 : Weintraub

Vector species	Disease Association/ Phytoplasma group	Host Plants	Distribution
<i>Cacopsylla pyrisuga</i> (Forster)	Pear decline/16SrX-C (36)	Pears	Russia

Renvoi vers cette référence bibliographique :

36. Grbic V. 1974. Some injurious species of the family Psyllidae in pear orchards in Vojvodina. *Zastita Bilja* 25:121--31 (article indisponible)

2009 : Hodkinson

Revue et synthèse sur la variation du cycle de vie

Table 1. (Continued.)

Higher taxon	Species	Climate zone	Plant type	Overwintering stage	Overwintering on host	Overwintering elsewhere	Volinism	Feeding site	Gall	Lerp former	Host plant(s)	References
Psyllinae	<i>C. pyrisuga</i> (Foerster)	TeM/M	Pd	A		C	1	S			<i>Pyrus</i>	Brocher (1926), Wojnarowska (1962), Lazarev (1975), Lauterer (1999), Ossiannilsson (1992)

Renvoi vers ces références bibliographiques :

Brocher F. 1926. Observations biologiques sur *Psylla pyrisuga* (Hemipt.). Ann Soc Entomol Fr. 95:183–188.

Lazarev MA. 1975. New data on the biology of the pear sucker *Psylla pyrisuga* (Homoptera: Psylloidea) in the Crimea. Entomol Obozr. 54:758–759.

Lauterer P. 1999. Results of the investigations on Hemiptera in Moravia, made by the Moravian Museum (Psylloidea 2). Cas Morav Mus, Brne. 84:71–151. (demande par mail le vendredi 17 mai – en attente)

Ossiannilsson F. 1992. The Psylloidea (Homoptera) of Fennoscandia and Denmark. Fauna Entomol Scand. 26:1–347.

2010 : Jarausch (publication non disponible)

In contrast to *C. pyri* and *C. pyricola*, the vector capability of **C. pyrisuga** is so far not yet confirmed (Jarausch and Jarausch 2010)

2012 : ARP Avis de l'Anses Saisine n° 2011-SA-0137

Les insectes-vecteurs sont le principal moyen de dissémination naturelle de cette maladie (Weintraub & Beanland, 2006). Il existe deux vecteurs essentiellement responsables de la dispersion de 'Ca. P. pyri': les psylles *Cacopsylla pyri* L. et *C. pyricola* F. (Garcia-Chapa et al., 2005 ; Seemüller & Harries, 2010). *C. pyri* est l'espèce la plus commune en Europe, *C. pyricola* étant moins fréquent (Siriez, 1982). *C. pyrisuga* est également mentionné comme autre insecte-vecteur potentiel de 'Ca. P. pyri' mais sa transmissibilité n'a pas encore été clairement démontrée (Jarausch & Jarausch, 2010).

Jarausch & Jarausch, 2010 : Jarausch B, Jarausch W (2010) Psyllid vectors and their control, In: Weintraub PG and Jones P (eds) *Phytoplasmas: Genomes, Plants Hosts and Vectors*, CAB International, Wallingford, UK, p. 250-270.

2014 : Jarausch (Review insect vectors Book COST)

“Seven distinct west Palaearctic Pyrus-feeding psyllid species have been described by Burckhardt and Hodkinson (1986). They considered this group as problematic: “(it) is a complex of related but distinct species with overlapping geographical distributions. Entomologists have usually applied names pyri or pyricola to all members of the complex and have failed to recognize morphological and biological differences between the species. Therefore, the literature concerning these species should be considered with a particular attention. A pictorial key supplemented with a dichotomous key is now available on web for the identification of the central European psyllids of the genus *Cacopsylla* which develop on Rosaceae, in particular Pyrus-feeding psyllid species, that should minimize the risk of misidentification” (<http://www.psyllidkey.com>). *Cacopsylla pyri* (Linné, 1758) and *Cacopsylla pyricola* (Foerster, 1848) have been described as vectors of 'Ca. P. pyri'. *C. pyri* is reported from Europe, the Caucasus, Central Asia, the Russian far east and China, while *C. pyricola* naturally occurs in the western palaeartics and has been introduced into the USA and Canada in the early 19th century. The two species are oligophagous on Pyrus species such as *P. communis*, *P. eleagrifolia*, *P. pyraster*, *P. amygdaliformis* and *P. salicifolia* where they produce several generations per year, while ***C. pyrisuga* is univoltine**; the adults overwinter on conifers and remigrate to Pyrus by middle March to April. *C. pyricola* has been described as vector of 'Ca. P. pyri' only for Great Britain (Davies et al., 1992) and North America (Jensen et al., 1964), while *C. pyri* was described as the vector in France (Lemoine, 1984), Italy (Carraro et al., 1998a) and Spain (Garcia-Chapa et al., 2005). Presentations during the COST meetings showed that individuals of *C. pyricola* and ***Cacopsylla pyrisuga* (Foerster, 1848)** have been found naturally infected with 'Ca. P. pyri' (e.g. Austria, Italy, Czech Republic, Slovenia, Croatia, Hungary, Bosnia and Herzegovina, Poland, Romania, Slovenia, and the Netherlands), **but their vector capability has not yet been proven by transmission trials.**”

2014 Torres (COST FA0807_Lisbon_final_meeting)

Relate expériences terrain dans un GT européen ; tests morphologiques & moléculaires ;

Czech Republic (CR) AP and PD were sporadically monitored in *C. picta* (90 individuals tested/4 positive) and *C. pyri* (966/11), ***C. pyrisuga*** (47/1), *C. pyricola* (17/1) in apple and pear orchards in East Bohemia, respectively. Examination of hemipterans in Czech vineyards revealed presence of phytoplasmas belonging to subgroups 16SrI-B, 16SrI-C, 16SrI-F, 16SrIII-B, 16SrXII-A and, unexpectedly, in *Jassargus obtusivalvis* of phytoplasmas of the 16SrXIV group.

Czech Republic (CR)
 'Candidatus Phytoplasma prunorum' is a long-standing problem in the CR. Survey was done in apricot, peach and cherry orchards. Different methods of DNA extraction and PCR detection were evaluated, including comparison of symptom variability after inoculation in different Prunus spp.. Occurrence of 'Candidatus Phytoplasma pyri' (PD), 'Ca. P. mali' (AP) and 'Ca. P. asteris' (165H-B and 165H-C subgroup) was confirmed in orchards as well as wild growing trees. Extensive study of AP revealed prevalence of P1 166-225 profile, rpkA subgroup, subtypes AP-15 and AT-2. 'Ca. P. solani' (stolbur) was detected in many plant species and its epidemics have been recorded in tomato, pepper and edoily in South Moravia. Study of genetic variability of stolbur also by use of polymorphic gene vmp 1 in annual crops and wild plants revealed presence of 4 genetic variants. First mass occurrence of *Hyalothoe obsoletus* (vector of stolbur) in South Moravia since 1992 was recorded in 2010. Alarming information were obtained about the infectivity of overwintering *Cacopsylla pruni* and *C. picta* that in which phytoplasmas were detected in up to 23% of the samples and *Anonotagallia ribaudi* with 50% of stolbur positive individuals. *Tribolium pretense* and *Chenopodius album* were found to be new hosts for stolbur phytoplasma that was also found to be associated with 'bois noir' (BN) disease of grapevine. *Aporus officinalis* and *Pionotaenionotaen* were found to be hosts for 'Ca. P. asteris' (AY, 165H-B). Phytoplasmas belonging to 165H-C ribosomal subgroup were firstly reported in *Echinacea purpurea*, *Rhododendron hybridum*, *Ribes rubrum* and *Agium graveolens*; phytoplasma of subgroup 165H-B in *E. purpurea* and *Chenopodium album* and phytoplasmas of 165H-A subgroup in *Rhododendron hybridum*. The report about *Ulmus minor* affected by elm yellows (EY) in Moravia represents the northernmost occurrence of this phytoplasma on elm trees within Europe. AP and PD were sporadically mentioned in *C. picta* (95 individuals tested/4 positive) and *C. pyri* (96/11), *C. pyricola* (47/1), *C. pyricola* (17/1) in apple and pear orchards in East Bohemia, respectively. Examination of hemipterans in Czech vineyards revealed presence of phytoplasmas belonging to subgroups 165H-B, 165H-C, 165H-F, 165H-A and, unexpectedly, in *Issoragus obtusivalvis* of phytoplasmas of the 165HIV group.

Austria ... Expectedly, *C. pruni* was found positive for ESFY in molecular tests whilst the Pear decline phytoplasma (PD; 'Ca. P. pyri') was detected in *C. pyricola*, *C. pyri* and **C. pyrisuga**.

Austria
 Significant increase of BN infected vineyards has been noted in the recent years in Austria. Systematic field investigations of the diseased vineyards as pathosystems consisting of infected vines, insect vectors and weed phytoplasma enabled both detecting new members of the pathosystem and experimenting with measures like pruning and pollarding to manage BN. Agallinae leafhopper *Anonotagallia ribaudi* captured on grapevine, and reared in the laboratory, was proven to transmit BN to *Vicia faba* in the experimental conditions. The BN transmission by *A. ribaudi* was also demonstrated to occur from *Convolvulus orvensis* to *Catharanthus roseus* (Riedle-Bauer et al., 2013) introducing potentially a new member in the BN epidemiological cycle. In this recent grapevine epidemiology research, the evidence of nettle and bindweed infections with 'Ca. P. convolvuli' was found. The outbreaks of FD in Austrian vineyards were not recorded until 2009. The foci in Southeast Styria and later in South Styria regions included indigenous varieties and an American hybrid, as well as Clematis plants. Mixed infections of BN and "Flavescence dorée" phytoplasmas (FD) were also recorded. Apricot, pear and apple orchards were surveyed for psyllid vectors and the insects were investigated for phytoplasma presence. The expected *Cacopsylla* species were found in the corresponding orchards but apparently *C. pruni* as an ESFY agent ('Ca. P. prunorum') vector has become a major concern in the Austrian apricot production. Expectedly, *C. pruni* was found positive for ESFY in molecular tests whilst the pear decline phytoplasma (PD; 'Ca. P. pyri') was detected in *C. pyricola*, *C. pyri* and **C. pyrisuga**. In apples, both *C. melanoneura* and *C. picta* were present but without AP detected, as yet.

Croatia 'Ca. P. asteris' (AY) was sporadically detected in grapevines, *S. titanus*, fruit trees and *C. pyri*. Even though symptoms of PD and AP were reported in the country long ago, the molecular characterization of pathogens and the vector survey are more recent. 'Ca. P. pyri' and 'Ca. P. prunorum' are the most widespread and have the highest incidence in pears and several stone fruit species, respectively. *C. pruni* has been implicated in the ESFY transmission, whilst *C. pyri* and **C. pyrisuga** were found positive for PD. Molecular diversity of fruit tree phytoplasmas was richer than expected which was later confirmed by multigene sequence analysis of some local strains within an international consortium. Interestingly, BN was found in pears as often as PD. Moreover, some **C. pyrisuga** samples also harboured stolbur phytoplasmas. The epidemiological significance of these findings is still unresolved.

infos aussi disponible sur le site web COST GT2 dédié

<http://www.costphytoplasma.ipwgnnet.org/WG2.htm> en particulier dans le rapport des expériences

<http://www.costphytoplasma.ipwgnnet.org/WG2/Phytoplasma%20Diseases%20and%20Vectors%20in%20Europe%20and%20Surroundings.docx>

COST Action FA0807								
WG 2/Task 2 QUESTIONNAIRE								
Title: Integrated Management of Phytoplasma Epidemics in Different Crop Systems								
Working Group 2: Epidemiology and Vector Ecology								
Task 2: MONITORING THE PRESENCE OF PHYTOPLASMA DISEASES AND THEIR PUTATIVE VECTORS IN DEFINED REGIONS THROUGH EUROPE								
Your address								
Name	Institution country/region			email / telephone				
Barbara Jarausch	RLP AgroScience, AlPlanta_Institute for Plant Research, Neustadt a.d.Weinstrasse, Rhineland-Palatinate, Germany			barbara.jarausch@agroscience.rlp.de ++49-6321-6711308				
Vectors (acknowledged)	Vectors (putative)	Population density	Phytoplasma/s	Infection rate	Detection methods	Vector host/s	Collection method/s	Identification method/s

not yet proven	Cacopsylla pyri C. pyricola C. pyrisuga	High low low	Ca. P. pyri	??		Pyrus	Beat tray	morphological
----------------	---	------------------------	-------------	----	--	-------	-----------	---------------

Phytoplasma vectors (putative/acknowledged) identified in your country/region								
Vectors (acknowledged)	Vectors (putative)	Population density #	Phytoplasmas	Infection rate	Detection method/s	Vector hosts	Collection method/s	Identification method/s

Austria

	<i>Cacopsylla pyrisuga</i>	moderate	'Candidatus Phytoplasma pyri'	unknown	PCR, RFLP	<i>Pyrus communis</i>	beating tray method	morphological identification
--	----------------------------	----------	-------------------------------	---------	-----------	-----------------------	---------------------	------------------------------

Croatie

	<i>Cacopsylla pyrisuga</i>	It was present	PD, BN	Number of	Molecular	Pear	entomological	morphological
	<i>pyrisuga</i>	in low-number populations		tested individuals is not enough to determine the infection rate. Few insects tested positive.	(PCR/RFLP)		net	

2017 Cho

En Corée et Chine, révision de la taxonomie sur la base morphologique ; 4 espèces sur le poirier, erreur d'identification de *C. pyrisuga* car deux formes suivant la saison ; rangement par groupe

The confused taxonomy of the east Palearctic pear psyllids, serious pests on cultivated pear, is reviewed. Fifty-six nominal species have been reported from *Pyrus*, 25 of which we consider valid and ten as not being associated with *Pyrus*. Our taxonomic revision suggests that, in Korea, **four *Cacopsylla* species develop on pear**: the univoltine *C. burckhardti* Luo et al. previously **misidentified** as *C. pyrisuga* (Foerster), the polyvoltine, seasonally dimorphic *C. jukyungi* (Kwon) (winter form 'cinereosignata' Luo et al., summer form 'jukyungi'), commonly found in Korean pear orchards, and *C. maculatili* Li (winter form 'maculatili', summer form 'qiuzili' Li) previously misidentified as *C. pyricola* (Foerster) by some authors, as well as the probably polyvoltine but not dimorphic *C. sandolbaea* (Park & Lee). The former three species (*C. burckhardti*, *C. jukyungi*, misidentified as *C. chinensis* (Yang & Li), and *C. maculatili*) occur also in Japan. Keys to the adult and fifth instar immatures as well as short biological notes are provided, and *C. jukyungi* and *C. sandolbaea* are redescribed.

Psyllidae, *Cacopsylla*, taxonomy, host plant, *Pyrus*, Rosaceae, Manchurian pear, Asian pear, European pear, China, Japan, South Korea, Russian Far East

TABLE 1. Pear psyllids with synonymy and information on distribution, host species and pathogens. Plant species which are not confirmed hosts by the presence of immatures are marked with 'nc'.

Name	Synonyms	Distribution	Host species	Vector
Cacopsylla pyrisuga-group				
<i>Cacopsylla burckhardtii</i> Luo et al., 2012		China (Gansu), Japan, Korea, Russia (Far East) (Luo et al. 2012; present paper)	<i>P. pyriformis</i> , <i>P. ussuriensis</i>	
<i>Cacopsylla melanonera</i> (Foerster, 1848)	<i>Psylla pisyophila</i> Flor, 1861; <i>Psylla oxyacanthiae</i> Meyer-Dür, 1871; <i>Psylla similis</i> Meyer-Dür, 1871	Algeria, Armenia, Austria, Azerbaijan, Belgium, Bulgaria, Croatia, Czech Republic, Denmark, France, Georgia, Germany, Greece, Hungary, Ireland, Italy, Lithuania, Moldova, Netherlands, Norway, Poland, Romania, Russia, Slovenia, Slovakia, Spain, Sweden, Switzerland, Tajikistan, Ukraine, UK, Uzbekistan	<i>Crataegus</i> spp.; also on <i>Malus domestica</i> , <i>Meispilus germanica</i> , <i>Pyrus communis</i>	on <i>Malus domestica</i> : <i>Candidatus</i> Phytoplasma mali, <i>Candidatus</i> Liberibacter europaeus
<i>Cacopsylla pyrisuga</i> (Foerster, 1848)	<i>Psylla austriaca</i> Flor, 1861; <i>Psylla aurantiaca</i> Goureau, 1861; <i>Psylla rufitarsis</i> Meyer-Dür, 1871; <i>Psylla rutila</i> Meyer-Dür, 1871	Armenia, Austria, Azerbaijan, Belarus, Czech Republic, Denmark, France, Georgia, Germany, Greece, Hungary, Iran, Israel, Italy, Lebanon, Norway, Serbia, Slovenia, Spain, Sweden, Switzerland, Turkey, UK	<i>P. amygdaliformis</i> , <i>P. communis</i> , <i>P. pashia</i> , <i>P. salicifolia</i>	<i>Candidatus</i> Phytoplasma pyri, <i>Candidatus</i> Liberibacter europaeus
<i>Cacopsylla accincta</i> Luo et al., 2012		China (Gansu)	<i>P. ussuriensis</i> (nc)	
Cacopsylla pyri-group				
<i>Cacopsylla betulaeifoliae</i> (Yang & Li, 1981)	<i>Psylla heterobetulaeifoliae</i> Yang & Li, 1981, <i>syn. nov.</i>	China (Anhui, Beijing, Hebei)	<i>P. betulifolia</i>	
<i>Cacopsylla bidens</i> (Sulc, 1907)	<i>Psylla vassilevi</i> Sulc, 1915; <i>Psylla jiangli</i> Yang & Li, 1981, <i>syn. nov.</i>	Armenia, Bulgaria, China (Xinjiang), Egypt, Estonia, France, Greece, India, Iran, Israel, Italy, Jordan, Kazakhstan, Kyrgyzstan, Latvia, Lebanon, Moldova, Mongolia, Romania, Slovakia, Slovenia, Turkmenistan, Ukraine, Uzbekistan; introduced into Argentina, Chile, Uruguay (Valle et al. 2017)	<i>P. communis</i> , <i>P. pyraster</i> , <i>P. syriaca</i>	
<i>Cacopsylla chinensis</i> (Yang & Li, 1981)	<i>Cacopsylla guangdongli</i> Li, 1993	China (Anhui, Beijing, Gansu, Guangdong, Guizhou, Hebei, Hubei, Jilin, Liaoning, Nei Mongol, Ningxia, Quizhou, Shaanxi, Shanxi, Shandong, Xinjiang, Zhejiang), Taiwan	<i>P. bretschneideri</i> , <i>P. pyriformis</i> , <i>P. ussuriensis</i> ; rarely on <i>P. communis</i>	PDTW phytoplasma
<i>Cacopsylla donggangica</i> Li, 2011		China (Gansu) (Li 2011)	<i>Pyrus</i> sp. (nc)	
<i>Cacopsylla fera</i> (Baeva, 1968)		Tatkhikistan, Uzbekistan	<i>P. bucharica</i>	
<i>Cacopsylla flavianthracina</i> Li, 2011		China (Anhui)	<i>Pyrus</i> sp. (nc)	
<i>Cacopsylla gyrogenna</i> Li, 2011		China (Shaanxi)	<i>Pyrus</i> sp. (nc)	

.....continued on the next page

2019 Jarausch

Cacopsylla Pyrisuga : vecteur présumé

Table 3.1 Phytoplasma diseases of fruit crops, their associated agents, psyllid vectors and vector's host plant

Psyllid species	Phytoplasma	Disease	Host plant
<i>Cacopsylla pyrisuga</i> ^b	' <i>Candidatus</i> Phytoplasma pyri'	Pear decline	<i>Pyrus</i> spp.

Phytoplasma vector species on apple and stone fruits are – like other North temperate psyllids – univoltine whereas pear psyllids with the exception of **Cacopsylla pyrisuga** (Foerster 1848) are mostly polyvoltine with overlapping generations.

In Europe, three recognized or presumed vectors of Pear decline disease (PD) live on pear: *Cacopsylla pyri* (Linné 1758) (Fig. 3.5), *C. pyricola* (Foerster 1848), and *C. pyrisuga* (Foerster 1848).

In contrast, **C. pyrisuga** is univoltine; the adults overwinter on conifers and remigrate to *Pyrus* spp. by middle March to April. Egg deposition takes place in two different steps at the beginning of April and second in the middle of May followed by a 6 week lasting nymph development and the emergence of new adults in June. All three pear psyllids can cause direct damage on pear trees: the nymphs affect plant growth by sucking phloem-sap, while the secreted honeydew burns plant tissue and favours the growth of sooty mold.

In contrast to *C. pyri* and *C. pyricola*, the vector capability of ***C. pyrisuga*** is so far not yet confirmed (Jarusch and Jarusch 2010).

2021 Monero

Extrapolation de *C. pyrisuga* à l'auteur Jensen de 1964 fausse

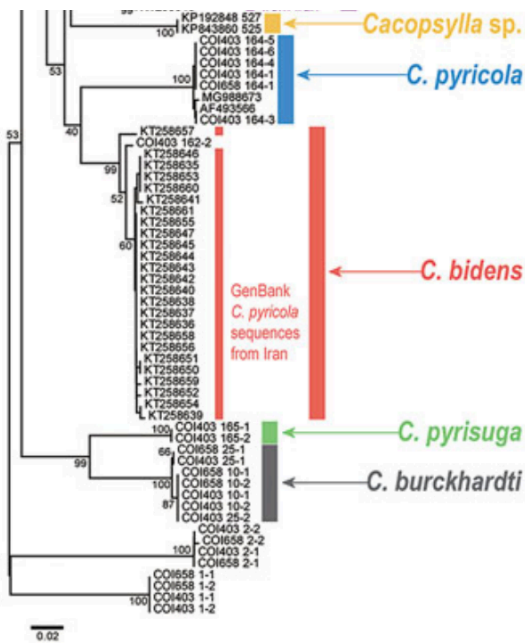
Cacopsylla pyri, *C. pyricola*, and *C. pyrisuga* (Foerster 1848; Hemiptera: Psyllidae) are the vectors of *Ca. P. pyri*, the causal agent PD, initially considered a virus (Jensen et al. 1964). *Cacopsylla pyri* is reported from Europe, the

Cacopsylla pyri, *C. pyricola*, and *C. pyrisuga* (Foerster 1848; Hemiptera: Psyllidae) are the vectors of *Ca. P. pyri*, the causal agent PD, initially considered a virus (Jensen et al. 1964).

2020 Cho

En Asie, révision de la taxonomie à partir d'analyse génétique ; Cho et al. (2020) used DNA barcoding

C. pyricola and ***C. pyrisuga*** from East Asia = misidentification of *C. jukyungi* and ***C. burckhardti***, respectively



2022 Riedle-Bauer

Vector transmission and epidemiology of 'Candidatus Phytoplasma pyri' in Austria and identification of *Cacopsylla pyrisuga* as new pathogen vector

Monika Riedle-Bauer¹ · Caroline Paleskić^{1,2} · Christina Schönhuber² · Martina Staples¹ · Günter Brader² 

Vector transmission and epidemiology of 'Candidatus Phytoplasma pyri' in Austria and identification of *Cacopsylla pyrisuga* as new pathogen vector

“ **Pear decline**, induced by the phytoplasma '**Candidatus Phytoplasma pyri**', transmitted by pear psyllids, is one of the most devastating diseases on *Pyrus communis* in Europe and North America. Investigations of pear psyllids in 4 pear orchards in lower Austria showed the presence of *Cacopsylla pyri*, *C. pyricola* and *C. pyrisuga* at all locations. PCR analyses revealed overall phytoplasma infection rates for *C. pyri* of 5.4%, for *C. pyricola*, of 4.6%, for *C. pyrisuga* remigrants of 9.6% and **for *C. pyrisuga* emigrants of 0%**. The rates of PCR-positive *C. pyri* and *C. pyricola* individuals varied greatly in the course of the year, and the highest infection rates were observed in late summer, autumn and in late winter. In **transmission experiments** with healthy pear seedlings, winterform individuals of *C. pyri* and *C. pyricola* transmitted the pathogen to 19.2% (5 out of 26) and 4.8% (2 out of 41) of the test plants, respectively. **The vectoring ability of *C. pyrisuga* was experimentally proven for the first time, and in transmission experiments with remigrants, 9.5% (2 out of 21) of the pear seedlings were infected.** Our data indicate a significant risk of pathogen transmission in pear orchards during the greater part of the year, especially in late winter, early spring and autumn. Multilocus sequence analysis by aid of the genes *aceF* and *imp* allowed the discrimination between 15 phytoplasma types. Three so far undescribed *aceF* genotypes and four undescribed *imp* genotypes were identified.”

2023 Civolani

Revue ; selon l'auteur, le travail taxonomique doit être ré-évalué avec citation d'auteurs à titre d'exemples ;

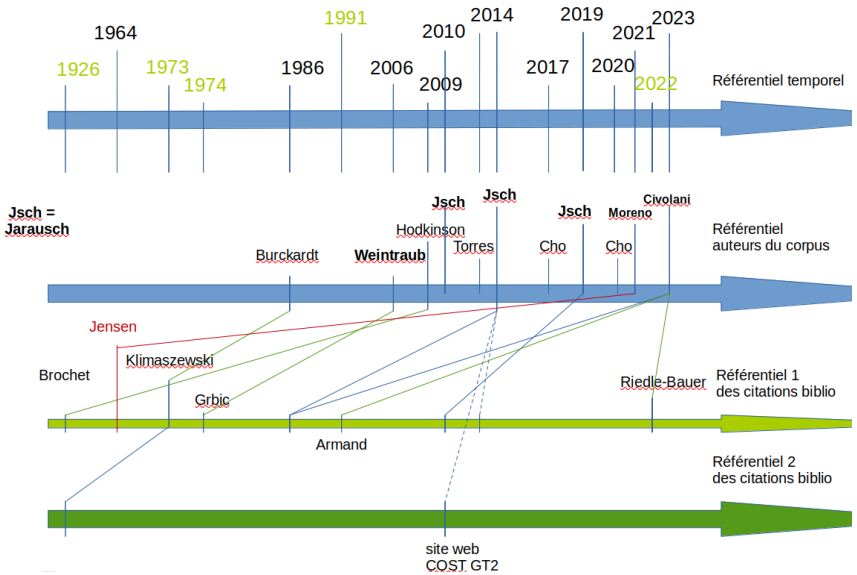
“Taxonomic work has begun to address the confusing group of pear psyllids in the Eastern Palaearctic Region, encompassing areas of eastern Russia, China, Taiwan, Japan, and Korea (Yang and Huang 2004, Luo et al. 2012, Cho et al. 2017, 2020, Chen et al.2018). Species in the Eastern Palaearctic region commonly associate with the Asian pears *P. pyrifolia* and *P. ussuriensis* (Cho et al.2017). **Analysis of this fauna has relied on use of molecular genetics, morphology, geography, and host information.** This approach has been used to untangle difficult complexes and to correct historical mistakes in species assignments (Luo et al. 2012, Cho et al. 2017, 2020, Chen et al. 2018). Cho et al. (2020) used DNA barcoding to separate species and to correct mistakes in the stated distributions of 2 Western Palaearctic species, *C. pyricola* and *C. pyrisuga*. These species at one time were thought to occur in East Asia but are now known to have been misidentifications of east Palaearctic species (Cho et al. 2017, 2020). There is no evidence for presence of any European pear psyllid in Eastern Asia (Cho et al. 2020). The synthesis of Cho et al. (2017) includes aspects of geography, life cycles, and host associations for the East Palaearctic pear psyllids, and provides keys to the Japanese and Korean species.”

“Known vectors of *Phytoplasma pyri* are *C. pyricola*, *C. pyri*, and ***C. pyrisuga*** (Jensen et al. 1964, Lemoine 1991, **Riedle-Bauer et al. 2022**).”

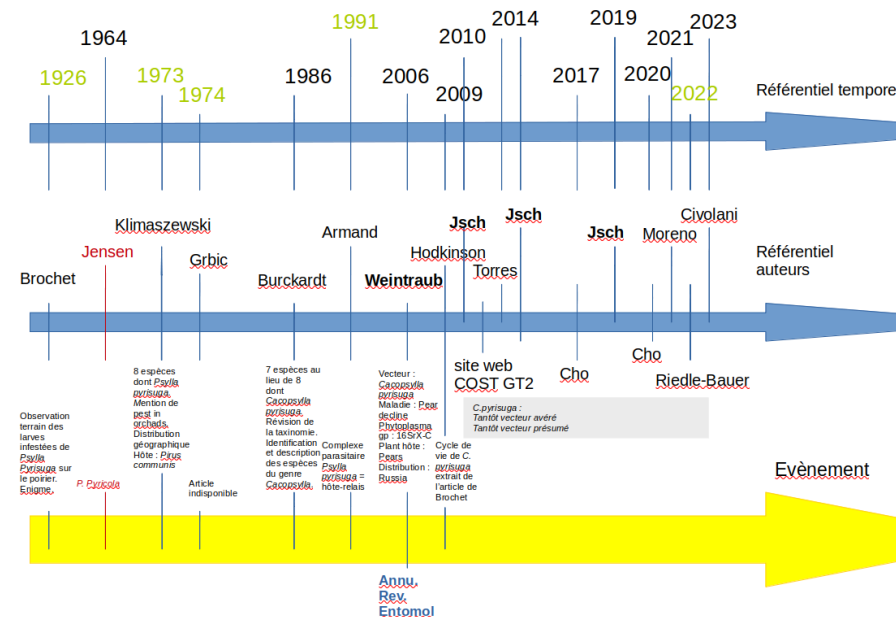
Référence : Riedle-Bauer M, Paleskić C, Schönhuber C, Staples M, Brader G. Vector transmission and epidemiology of “Candidatus Phytoplasma pyri” in Austria and identification of Cacopsylla pyrisuga as new pathogen vector. J Plant Dis Prot. 2022;129(2):375–386. <https://doi.org/10.1007/s41348-021-00526-y>

Extraction de la Bibliographie :

Armand E, Lyoussofi A, Rieux R. Évolution du complexe parasitaire des psylles du poirier *Psylla pyri* et ***Psylla pyrisuga*** (Homoptera: Psyllidae) en vergers dans le sud-est France au cours de la période hivernale, printanière et estivale. Entomophaga. 1991;36(2):287–294. <https://doi.org/10.1007/Bf02374564>



Les dates en noir reflètent les publications scientifiques disponibles dans le corpus initial. Les dates en vert représentent les publications citées extraites des publications du corpus initial, mais non disponibles dans le corpus initial.



Classification

Rechercher une espèce

Règne: **Animalia**

Phylum: **Arthropoda**

Classe: **Insecta**

Ordre: **Hemiptera**

Famille: **Psyllidae**

Genre: **Cacopsylla Ossiannilsson, 1970**

Espèce: **Cacopsylla pyrisuga (Foerster, 1848)**

- *Psylla aurantiaca* Goureau, 1861
- *Psylla austriaca* Flor, 1861
- *Psylla pyrisuga* Foerster
- *Psylla rufitarsis* Meyer-Dür, 1871
- *Psylla rutila* Meyer-Dür, 1871

Aucun enfant

ESPECE | ACCÉPTÉ

Cacopsylla pyrisuga (Foerster, 1848)

Publié dans: Goureau C. Les Insectes nuisibles aux arbres fruitiers, aux plantes potagères, aux céréales et aux plantes fourragères. 1: 366 pp. V. Masson et fils, Paris (France), (1861).
source: Psylloidea database

Basionyme: *Psylla pyrisuga* Foerster

181 OCCURRENCES

VUE D'ENSEMBLE 2 TRAITEMENTS TAXONOMIQUES STATISTIQUES TAXON DE RÉFÉRENCE ⇐

21 OCCURRENCES AVEC DES IMAGES

24 ENREGISTREMENTS GÉORÉFÉRÉS

EXPLORER

<https://www.gbif.org/fr/species/2012868>

ANNEXE n°11 - Tableau d'analyse de quelques indices linguistiques

EXEMPLE PHRASE	TYPE OF EVENT	EVEN STATUS
Psyllids as vectors of bacteria in plant health: a case study of psyllids vectors of a phytoplasma on pear	psylle occurrence + ~ life-cycle	confirmed (affirmation ; titre)
Phytopathogenic agents can be transmitted either through seeds, pollen, vegetative propagation, grafting, or via mobile organisms called vectors.	"can be transmitted ... via mobile organisms called vectors"	hypothetic
These vectors enable the pathogen to be disseminated over varying distances and to penetrate the tissues of their host plant (Herrbach et al., 2013)	"These vectors enable the pathogen to be disseminated over varying distances and to penetrate the tissues of their host plant" = phytoplasma transmission to plant	confirmed ("enable" - présent)
Pear decline is caused by the phytoplasma 'Candidatus Phytoplasma pyri' (CaPpyri) (Seemüller and Schneider, 2004).	diseased plant	confirmed (is caused by)
This disease, known as Pear decline (PD) (Civolanie et al., 2023), is caused by the phytoplasma 'Candidatus P.pyri' (Seemüller and Schneider, 2004).	diseased plant	confirmed (is caused by)
Experiments in the state of Washington implicated the pear psylla (<i>Psylla pyricola</i> Foerster) , as causing the disease by introducing a nonviral insect toxin into the pear trees while feeding. Research in California also indicated that the pear psylla played a key role in the problem , but some of the results could be better attributed to a psylla-transmitted virus than psylla toxin alone. The 1963 entomological investigations carried out in California on the relation of pear psylla to Pear decline, now provide very strong evidence that quick decline disease of pears is due to a virus transmitted by	diseased plant	confirmed ("provide very strong evidence")

EXEMPLE PHRASE	TYPE OF EVENT	EVEN STATUS
<p>the pear psylla rather than by an insect toxin alone. The transmission experiments reported here were carried out in orchard plots at Davis. The test trees were commercial varieties, mostly Bartlett, on <i>Pyrus serotina</i>, Rehder (Oriental type) seedling rootstocks.</p>		
<p>Part of the test trees in both plots were made available for six to eight days for feeding by adult psylla that had been transferred from slow-decline pear trees for virus transmission experiments.</p>	<p>phytoplasme presence in psylle</p>	<p>confirmed</p>
<p>Approximately two months after the test feedings began, quick decline symptoms (wilting and collapse of foliage and succulent shoots) developed in several of the trees that had been exposed to feeding for only six to eight days by adult psylla transferred from slow decline trees.</p>	<p>diseased plant (symptômes de déclin mais est-ce vraiment le Pear decline ?)</p>	
<p>The results obtained in these experiments indicate that psylla-transmitted virus causes the sudden wilting and collapse of pear trees and that psylla toxin – even in the relatively large and continuous doses over a period of three months – does not cause quick decline in vigorously growing trees, in the absence of the virus.</p>	<p>diseased plant</p>	
<p>The present findings do indicate, however, that a psylla-transmitted virus is responsible for most of the tree losses that have been associated with Pear decline disease.</p>	<p>diseased plant</p>	<p>confirmed</p>
<p>The ‘life cycle’ of Pear decline phytoplasma in the vector <i>Cacopsylla pyri</i> has been studied. For two consecutive years, groups of insects caught monthly from March to October were used for transmission trials. The psylla was equally infective throughout the experimental period. In the second year, insects were fed on dormant plants from November to March and regularly tested. Retention of infectivity of the Pear decline phy-toplasma in the vector lasted through the winter. The overwintering adults of <i>C. pyri</i> did not transmit the phytoplasmas to plants in dormancy, but they did, without any previous recharge feeding, after bud break in the spring. Control of the overwintering adults of <i>C.pyri</i></p>	<p>no transmission to plant in dormancy ("did not transmit") from November to March</p>	<p>confirmed</p>

EXEMPLE PHRASE	TYPE OF EVENT	EVEN STATUS
appears to be of fundamental importance for preventing the spread of Pear decline.		
Pear decline (PD) is one of the most dangerous dis-eases of pear trees. The causal agent is a phytoplasma belonging to the apple proliferation group (Seemüller et al., 1998). In the late 1950s, PD affected more than a million pear trees along the Pacific Coast of North America (Woodbridge et al., 1957). The disease has long been known as 'moria' in northern Italy (Refatti, 1948) and has been reported in other European and extra-European fruit areas (Blattny and Vana, 1974; Davies et al., 1992; Lorenz et al., 1995; Jarausch and Dosba, 1995; Malinowski et al., 1996; Battle et al., 1999). Recently, PD has also been detected in southern Italy, in areas previously considered free from the disease (Marcone et al., 1996).	diseased plant	nouvel événement : 'Recently, PD has also been detected in ... in areas previously considered free from the disease (Marcone et al., 1996).'
Jensen et al. (1964) showed that pear psylla (Psylla pyri-cola Förster, now Cacopsylla pyricola Förster) transmits 'a virus' capable of causing PD disease.	phytoplasme transmission to plant ; diseased plant	confirmed ('capable of causing')
Our results confirm that C. pyri is a particularly active vector of the PD agent. Its presence was detected in 35 out of 64 (55%) groups of 25 psyllas collected in orchards. In terms of single insects, the minimum percentage of psyllids carrying phytoplasmas was 2.2% during the whole experimental period and, in particular, 1.60% from March to June and 2.75% from July to October. Likewise, 17 out of 56 (30%) of the inoculated test plants during the vegetative phase (using a total of about 3000 insects) became infected. These results explain the rapid spread of Pear decline in Italy, where C. pyri is the predominant species in pear orchards. ...	phytoplasme presence in psylle	

EXEMPLE PHRASE	TYPE OF EVENT	EVEN STATUS
<p>Table 1 shows that the transmission efficiency of the groups of psylla gathered during July-October was higher (69%) than that for March-June (41%). Our trials also showed that C. pyri retained infectivity during winter but could not transmit PD to dormant test plants. However, the following spring the same overwintering adults could transmit the phytoplasma to test plants, without recharge feeding. The lack of transmission may be because pear trees are not infectable at this time and/or because the vector is unable to inoculate the host during dormancy. All plants colonized by infected psyllas in winter during the retention experiment were negative by PCR-RFLP. It is clear that such plants can be considered as 'hosts for the vector but completely resistant to the PD agent'; therefore they could be ideal for investigations on retention of infectivity in a potential vector.</p>	<p>no transmission to plant in dormancy</p>	<p>confirmed : could not transmit PD... could transmit ...</p>
<p>Both, phytoplasmas and psyllid vectors, are geographically limited to the Palaearctic region, mainly Europe.</p>	<p><i>relation spécifique</i></p>	
<p>Thus, detailed knowledge about the biology and ecology of the vector species, their host plants as well as knowledge about the transmission parameters is crucial.</p>	<p><i>"knowledge about the transmission parameters is crucial" = transmission</i></p>	
<p>Table 3.1 gives an introductive over-view of the most important phytoplasma diseases and their agents which are transmitted by psyllid vectors.</p>	<p><i>relation spécifique</i></p>	
<p>Table 3.1 Phytoplasma diseases of fruit crops, their associated agents, psyllid vectors and vector's host plant</p>	<p><i>relation spécifique</i></p>	
<p>So far, all important fruit crop diseases associated with the presence of phytoplasmas of the group 16SrX are vectored by psyllids.</p>	<p><i>relation spécifique</i></p>	

EXEMPLE PHRASE	TYPE OF EVENT	EVEN STATUS
In Europe, three recognized or presumed vectors of Pear decline disease (PD) live on pear: <i>Cacopsylla pyri</i> (Linné 1758) (Fig. 3.5), <i>C. pyricola</i> (Foerster 1848), and <i>C. pyrisuga</i> (Foerster 1848).	diseased plant	
First reports of pear psyllids as vectors of phytoplasmas came from the Pacific coast of North America. Jensen et al. (1964) identified <i>C. pyricola</i> as the vector of 'Ca. P. pyri' at a time when the disease was thought to be virus-borne. Since then no further vector has been described for the USA.	psylle occurrence	unconfirmed
However, the distribution of the putative vectors of 'Ca. P. pyri' in Europe and the whole Palearctic region is diverse: while for Great Britain only <i>C. pyricola</i> has been described as vector (Davies et al. 1992), <i>C. pyri</i> was identified as main vector in France (Lemoine 1984), Italy (Carraro et al. 1998) and Spain (Garcia-Chapa et al. 2005).	psylle occurrence	confirmed
In Czech Republic (Kucerova et al. 2007; Ludvikova et al. 2011) and in lower Austria (Lethmayer et al. 2011) all three pear psyllid species were found naturally infected with 'Ca. P. pyri', while in Turkey only <i>C. pyri</i> carried the phytoplasma albeit all the three species were present in the investigated regions (Kaya et al. 2016).	phytoplasme presence in psylle	confirmed
In contrast to <i>C. pyri</i> and <i>C. pyricola</i> , the vector capability of <i>C. pyrisuga</i> is so far not yet confirmed (Jarausch and Jarausch 2010).	psylle occurrence	unconfirmed (in contrast ... not yet confirmed)
Recently, a fourth pear psyllid, the polyvoltine species <i>Cacopsylla bidens</i> (Šulc 1907) has been found infected with 'Ca. P. pyri' in Bulgaria (Etropolska et al. 2015).	phytoplasme presence in psylle	confirmed
Also for this species its vector capability still needs to be proven.	psylle occurrence	hypothetic ("still needs to be proven")

EXEMPLE PHRASE	TYPE OF EVENT	EVEN STATUS
<p>After the identification of <i>C. pyricola</i> as vector for 'Ca. <i>P. pyri</i>' in California (Jensen et al. 1964), many investigations in the USA and Europe followed in order to determine the infection rate of the psyllids and to analyse the transmission param-eters.</p>	<p>determine the infection rate of the psyllids and to analyse the transmission param-eters'</p>	
<p>In United Kindgdom transmission trials carried out with field-collected <i>C. pyricola</i> yielded transmission rates between 3–61% depending on the collection site of the psyllids (Davies et al. 1992).</p>	<p>transmission trials'</p>	
<p>Acquisition of 'Ca. <i>P. pyri</i>' by <i>C. pyricola</i> from experimentally infected pear seedlings was best in August and lowest in winter.</p>	<p>phytoplasma presence in psylle</p>	<p>confirmed</p>
<p>In California, Blomquist and Kirkpatrick (2002a) detected the pathogen in both winter and summer forms of <i>C. pyricola</i>, but without a clear seasonal trend.</p>	<p>phytoplasme presence in psylle</p>	
<p>They concluded that psyllid-mediated spring infec-tions could happen well before 'Ca. <i>P. pyri</i>' would normally recolonize the upper part of the tree from the roots.</p>	<p>phytoplasma transmission to plant</p>	<p>hypothetic des 3 phrases avec modalités : 'They concluded that psyllid-mediated spring infec-tions could happen well before 'Ca. <i>P. pyri</i>' would normally recolonize the upper part of the tree from the roots.'</p>
<p>In Italy, Carraro et al. (1998, 2001a) detected 'Ca. <i>P. pyri</i>' in 55% of groups of <i>C. pyri</i> collected from March to October in the orchards and 30% of the inoculated test plants became infected.</p>	<p>phytoplasma transmission to plant</p>	

EXEMPLE PHRASE	TYPE OF EVENT	EVEN STATUS
They could furthermore show that <i>C. pyri</i> retained the phytoplasma during winter, but could not transmit PD to dormant plants.	phytoplasma transmission to plant	hypothetic des 3 phrases avec modalités : "They could furthermore show that <i>C. pyri</i> retained the phytoplasma during winter, but could not transmit PD to dormant plants."
Raddadi et al. (2011) tested <i>C. pyri</i> collected in Italian pear orchards detecting 'Ca. <i>P. pyri</i> ' in 26% of the specimens (in 27% of males, and in 25% of females) with nested-PCR and in 51% of the individuals through quantitative PCR with a number of 'Ca. <i>P. pyri</i> ' cells ranging from 1.43×10^1 to 8.50×10^5 per <i>C. pyri</i> individual.	phytoplasme presence in psylle	
Moreover, fluorescent in situ hybridization (FISH) allowed to detect 'Ca. <i>P. pyri</i> ' in Malpighian tubules and salivary glands of <i>C. pyri</i> .	phytoplasme presence in psylle	
Garcia- Chapa et al. (2005) found that the percentage of infected individuals in Spain is similar from June to August but reaching a rate of almost 100% in September coinciding with the maximum phytoplasma titre in the aerial plant parts.	phytoplasme presence in psylle	
Although the percentage of infected psyllids was similar for both genders, 'Ca. <i>P. pyri</i> ' transmission by females was significantly higher than by males.	phytoplasma transmission to plant	
During transmission trials under controlled conditions Caglayan et al. (2010) showed the capability of <i>C. pyri</i> to transmit PD from infected pears to healthy periwinkles and confirmed it as vector of 'Ca. <i>P. pyri</i> ' in Turkey.	During transmission trials under controlled conditions ; phytoplasma transmission to plant ; diseased plant	
PD has also been found in Taiwan (PD-TW) where the European species <i>C. pyri</i> and <i>C. pyricola</i> are not present. Liu et al. (2007) found two other <i>Cacopsylla</i> species, <i>C. qianli</i> and <i>C. chinensis</i> , infected with the PD-TW phytoplasma.	phytoplasma transmission to plant	

EXEMPLE PHRASE	TYPE OF EVENT	EVEN STATUS
Their role in trans-mission of PD-TW phytoplasma in Taiwan remains to be clarified.	phytoplasma transmission to plant	
But recently Liu et al. (2011) proved the transmission capacity of <i>C. chinensis</i> during transmission trials in Taiwan.	phytoplasma transmission to plant	
Based on PCR detection and symptom development they showed that pear trees were either infected by PD-TW or PD-TWII phytoplasma strains, or co-infected by both, when exposed to <i>C. chinensis</i> specimens.	diseased plant	
As two similar diseases, western X and PYLR, associated with two genetically distinct phytoplasmas exist in the same region, only the application of molecular methods enabled to proof that a psyllid is the main vector of PYLR phytoplasma.	psylle occurrence	hypothetic ("only the application of molecular methods enabled to proof that a psyllid is the main vector of PYLR phytoplasma")
Experimental transmission of PYLR phyto-plasma to peach seedlings was achieved with feld collected <i>C. pyricola</i> from natu-rally infected peach trees (Guerra 1997).	phytoplasma transmission to plant	(experiments ...was achieved)
Diversity, biology, and management of the pear psyllids: a global look	psylle occurrence	confirmed (affirmation ; titre)
Pear psyllids may cause any of several types of damage to commer-cial pear orchards, including russet and downgrading of fruit due to marking of the pear fruit by honeydew and sooty mold, prema-ture leaf drop and tree decline, and vectoring of the pathogen that causes "Pear decline".	diseased plant	hypothetic ("may cause any")

EXEMPLE PHRASE	TYPE OF EVENT	EVEN STATUS
One or more species of pear psyllid are found in all of these regions except for South Africa, Australia, and New Zealand (Cho et al. 2017, 2020).	psylle occurrence	confirmation (was found) ? ou hypothetic car citation ?
“Ca. Phytoplasma pyri” is the pathogen that causes decline dis-ease in pears, shown by yellowing of leaves, decreased tree vigor, and death of the tree.	diseased plant	
Known vectors of Phytoplasma pyri are C. pyricola, C. pyri, and C. pyrisuga (Jensen et al. 1964, Lemoine 1991, Riedle-Bauer et al. 2022).	phytoplasma presence in psylle	
The behavioral changes may improve vector competency.	phytoplasma presence in psylle (déduction liée aux phrases précédentes)	hypothetic ("may improve")
Little is known of whether Phytoplasma pyri affects behavior of the pear psyllids, although Cruz et al. (2018) showed that phytoplasma-infected C. pyricola were less likely than uninfected psyllids to disperse from the pear host and become captured on yellow traps.	phytoplasma presence in psylle	
First, while the most recent taxonomic overviews of the pear psyllids list 24 known species, much of what we know about the biology of the group is concentrated on a much smaller set of species consisting pri-marly of species having a west Palaeartic origin. Moreover, much of the research has focused on 2 European species, C. pyri and C. pyricola – the latter now a permanent and damaging component of North American pear orchards – and 1 species (C. bidens) found in a region encompassing areas of the Mediterranean basin, Eastern Europe, Western Asia, and South	presence of psylle	

EXEMPLE PHRASE	TYPE OF EVENT	EVEN STATUS
<p>America. Other than taxonomic study, much less is known of the remaining psyllid fauna, with the possible exception of the East Asia <i>C. chinensis</i>. Consequently, while our synthesis is stated to encompass a global overview of the pear psyllids, many species within the collection of 24 species have seen little or no biological research and consequently have received little attention in this review.</p>		
<p><i>Ca. pyri</i> (Linnaeus) ... Apple proliferation/16SrX-A (56) ... Apple ... Europe ... Pear decline/16SrX-C (9) ... Pear ... Europe</p>	<i>relation spécifique</i>	
<p><i>Ca. pyricola</i> (Forster) ... Pear decline/16SrX-C (48) ... <i>Pyrus</i> spp. ... North America</p>	<i>relation spécifique</i>	
<p><i>Cacopsylla pyrisuga</i> (Forster) ... Pear decline/16SrX-C (36) ... Pears ... Russia</p>	<i>relation spécifique</i>	