Characteristics of *Plasmodiophora brassicae* causing clubroot disease on the cruciferous weed *Cardamine occulta* in post-harvest paddy fields in Japan

（日本の収穫後水田においてアブラナ科雑草*Cardamine occulta*に根こぶ病を引きおこす*Plasmodiophora brassicae*の性状）

A THESIS
SUBMITTED TO THE UNITED GRADUATE SCHOOL OF AGRICULTURAL SCIENCES TOTTORI UNIVERSITY IN PARTIAL FULFILMENT OF THE REQUREMENTS FOR THE DEGREE OF DOCTORAL OF PHILOSOPY

BY
ANH TUNG PHAN LAM

2022

TABLE OF CONTENTS

CHAPTER 1
GENERAL INTRODUCTION OF CLUBROOT DISEASE
1 Background and Objectives 1
1.1.1 The important role of Brassica crops 1
1.1.2 The damage of clubroot disease to Brassica crops 1
1.1.3 The causal pathogen *Plasmodiophora brassicae* 2
1.1.4 The life cycle of *P. brassicae* 3
1.1.5 Symptomatology 6
1.1.6 Morphological and anatomical assessment 6
1.1.7 Histological assessment and microscopic examination 6
1.1.8 Genome of *P. brassicae* 7
1.1.9 The challenges of clubroot study 8
1.2.1 The cruciferous weed *Cardamine occulta* 10
1.2.2 Nomenclature 11
1.2.3 The polyploidy genome 12
1.2.4 The life cycle of C. occulta in rice paddy fields 13
1.3 Objectives 17

CHAPTER 2
MORPHOLOGICAL AND HISTOPATHOLOGICAL ASSESSMENT ON THE CLUBROOT DISEASE OF *C. OCCULTA*
2.1 Introduction 18
2.2 Materials and Methods 19
2.2.1 Field surveys on clubroot disease of *C. occulta* 19
2.2.2 Clubroot samples collected 20
2.2.3 Microscopic observation 20
2.24 Preparation of resin-embedded specimens and sections 20
2.3 Results 21

2.3.1 Characterization of Clubroot symptoms on C. occulta 21
2.3.2 Histological Analysis under Microscopic Investigations 25
2.4 Discussion 29
CHAPTER 3
GENETIC INVESTIGATIONS ON ITS REGIONS AND ACTIN GENES OF P. BRASSICAE REVEALING THE NOVEL POPULATIONS
3.1 Introduction 30
3.2 Materials and Methods 32
3.2.1 DNA extraction 32
3.2.2 Primer designation, Cloning and Sequencing 32
3.2.3 Phylogenetic analysis 33
3.2.4 The ITS2 secondary structure prediction 35
3.3 Results 35
3.3.1 Sequence Variation of *P. brassicae* 35
3.3.1.1 ActI gene sequencing analysis 35
3.3.1.2 ITS sequencing analysis 37
3.3.2 Phylogenetic analysis 39
3.3.3 ITS2 secondary structure analysis 41
3.4 Discussion 43
3.4.1 Coevolutionary relationship between *P.brassicae* and *C.occulta* 43
3.4.2 The crucial role of the novel group *P.brassicae*
to the clubroot epidemiology 45

CHAPTER 4
THE *P. BRASSICAE* POPULATIONS INFFECTED *C. OCCULTA* REVEAL THE COMPLEX DYNAMICS OF RIBOSOMAL INTRONS
4.1 Introduction 49
4.2 Materials and Methods 51
4.2.1 Clubroot samples collection 51
4.2.2 DNA extraction 51
4.2.3 Primer designation and sequencing analysis 51

4.2.4 Sequence Annotation 53
4.2.5 Secondary structure prediction 53
4.3 Results 53
4.3.1 The pattern of loss/gain introns distributed within
the *P. brassicae* populations in Japan 53
4.3.2 Intron analysis of Small (18S) and Large (28S) Subunit rRNA genes 55
4.3.2.1 Small subunit (18S) rRNA gene 55
4.3.2.2 Large subunit (28S) rRNA gene 58
4.4 Discussion 62
4.4.1 Early or late theory for the intronic distribution in *P. brassicae* 63
4.4.2 The intronic evolution equates to
speciation event of *P. brassicae* populations? 65
4.4.3 Prospective research for intragenic sequences of *P. brassicae* 67

CHAPTER 5

GENERAL DISCUSSION AND CONCLUSION 70
Supplementary Tables 72
Supplementary Figures 74
References 82
Summary 92
Japanese Summary 95
List of publications related to the thesis 98
Acknowledgement 99

CHAPTER 1

GENERAL INTRODUCTION OF CLUBROOT DISEASE

1. BACKGROUND AND OBJECTIVES

1.1.1 The important role of Brassica crops

Cruciferae or Brassicaceae family is one of the most diverse plant families due to the natural and artificial selection (Kagale et al. 2014). The genus Brassica is comprised of 54 genera and many of them are valuable crops (cole crops, cruciferous vegetables, ect…) that contribute economically and nutritionally to crop production worldwide (Rakow 2004, Wang et al. 2014). Under the domestication process, only one species *Brassica oleracea* has been bred and selected with many dominant cultivars such as cabbage, broccoli and cauliflower, kale…yet have been selected by mankind for several purposes of use, e.g. leaves, flowers, stems, and roots. In Korea, fermented cabbage Kimchi is a culture symbol. In Canada, canola oil is one of the most significant crops for oil industry. In Japan cabbage is a signature dish and was the second most consuming vegetable nationwide (Ministry of Agriculture, Forestry and Fisheries 2015). Other members of Brassicas also play important roles as supply for animal feed, food condiments, and ornamentals (Dixon 2006). Even the non-crop member such as the tiny weed *Arabidopsis thaliana*, has become the model organism for genetic research in plant since it was the first plant whose entire genome was sequenced (Anon 2000).

1.1.2 The damage of clubroot disease to the Brassica crops

Almost all continents except for Antarctica are threatened by clubroot disease (Fig.1). In the nutshell, wherever the Brassica crops are grown the clubroot shadow is present. This disease is prevalent from the East to the West

and cause a great damage to the Brassica crops up to 100% yield loss (Dixon 2009a, b).

Figure 1. Clubroot disease distribution on global scale, red dots illustrated the presence of clubroot. Image source retrieved from Saharan et al. (2021)

In China, clubroot disease is considered as a major cause of depressed yield with 17% loss of young growing plants, and 15% at maturity with 10.2% losses of yield. (Jing et al. 2008) In Canada, the canola oil industry was heavily affected by clubroot damage, 30-100% infection surveyed by Hwang et al. (2011). Clubroot also occurred in Australia where oilseed rapes yield reduced at least 50% in total (Donald and Porter 2003). Crete (1981) carried a global survey on clubroot infection that revealed high levels of infestation (>10%) were noted on *B. oleracea* in Australia, Canada, Czechoslovakia, Finland, Germany, Ireland, the Netherlands, Norway, Poland, Scotland, and Wales; in *B. rapa*, in Germany, New Zealand, and the USA; and in *B. napus*, in Finland, New Zealand, Scotland, and Wales. This survey has covered 6 million hectares of Brassica crops, and found an overall mean infection of 11%.

1.1.3 The causal pathogen *Plasmodiophora brassicae*

The causal pathogen of clubroot disease is a telluric and intracellular obligate protist, *Plasmodiophora brassicae*. The taxonomic status of this parasitic protist has been placed under the super group or the infra Kingdom "Rhizaria", within the Class "Phytomyxea" including in the Order "Plasmodiophorida" (Cavalier-Smith 1993, 2002, Goeche et al. 2012). Rhizaria is comprised of a diverse amoebae group known for their differentiated flagellae (Nikolaev et al. 2004). These amoeboid protists were misclassified as a fungus or a slime mold but were corrected to place in the kingdom Rhizaria by molecular phylogenetic analysis (Bass et al. 2005). According to Burki and Keeling (2014), the taxonomy of the *P. brassicae* is as under:

Kingdom: Rhizaria
Phylum: Cercozoa
Class: Phytomyxea
Order: Plasmodiophorida
Family: Plasmodiophoraceae
Genus: Plasmodiophora
Species: brassicae

1.1.4 The life cycle of *P. brassicae*

The parasitic protist *P. brassicae* has a complicated life cycle which still remain largely enigmatic to clubroot experts (briefly summarized in Fig.2). The full details of *P. brassicae* life cycle were meticulously described in several reports (Agrios 2005, Kageyama and Asano 2009, Schwelm et al. 2015). In summary, the *P. brassicae* spores rest in the soil for several decades until 20-30 years (Dixon 2009c). When crucifers are present, the resting spores stimulated by the root exudates from the cruciferous hosts, initiate the germination process. The spores release the primary zoospores which abruptly infect the root hairs system where they develop into sporangia. These sporangia continue to become the

secondary zoospores and release into the soil again. The first stage of infection was termed as the primary infection or the root-hair infection. At this stage, the infected plants usually go without symptoms, only visible under microscopic observation. The second stage begins where the secondary zoospores reinfect to the main root (tap root) system, specifically in the cortical cells of the main roots. They localize in the cortical tissues, where they undergo a series of cell divisions and eventually form multinucleated secondary plasmodia in the cortex and stele. The host cells are modified to go through extreme hyperplasia (cell division) and hypertrophy (cell expansion) to supply nutritious sources for the plasmodium clusters which later sporulate into spores, thus resting the soil. At this stage, symptoms can generally be observed by formed galls when uprooting the infected plants. Root galls are varied in shapes (clubbed-, spindled-, spheroid) and sizes (small, medium, large) depending on the severity degree. For severely infected cases, above-ground symptoms are observable with plant patches suffering wilting, stunting or yellowing (Dixon 1991, 2002). The second stage was termed as secondary infection or cortical infection. This stage only occurs when Brassica hosts are susceptible whereas in the resistant varieties, the zoospores fail to reinfect after the primary infection. The same also has happened to the non-host plants, the primary zoospores are not able to infect the main roots.

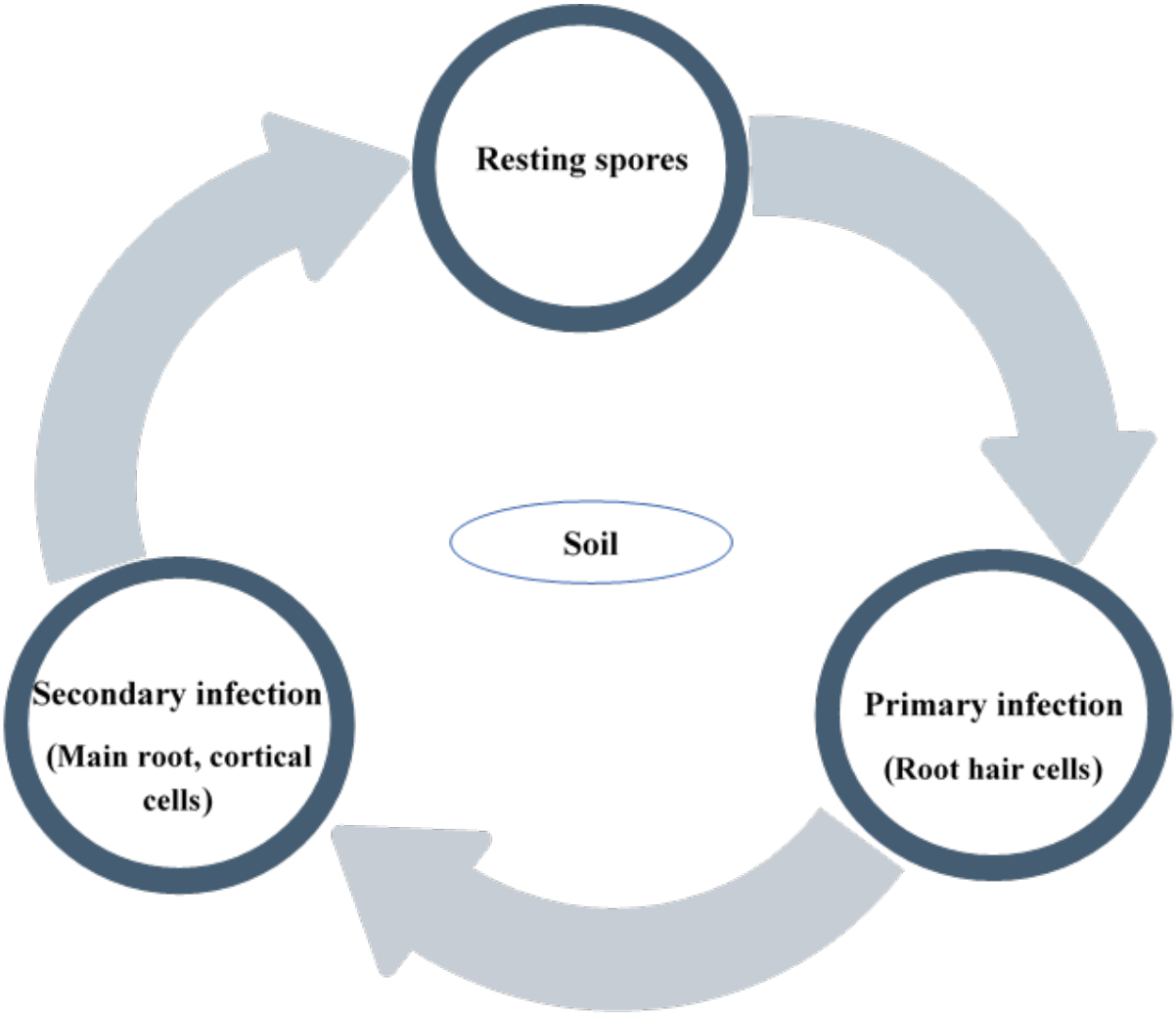


Figure 2. Simplified life cycle of *P. brassicae*. The graphic only summarized the three most significant stages through out the entire complicated life of *P. brassicae*.

1.1.5 Symptomatology

Clubroot symptoms very often goes undiagnosed when the above-ground symptoms are minor. In this case, infected plants are pulled up to have examined for galls and root appearances. For mild symptoms, many other diseases have similar characters, therefore the symptomology of clubroot disease should be verified with further investigations such as stained samples observed under the microscope or PCR-specific primer designed for *P. brassicae*. Morphological, anatomical and histological characterisation of clubroot disease have been described in details of several publications, books, clubroot-related websites….in this narrow scope, certain important features are selected to mention deliberately.

1.1.6 Morphological and anatomical assessment

Above-ground symptoms are similar to those of other common diseases diagnosed from pathogenic fungi or bacteria found commonly on Brassica crops such as Alternaria leaf spot (Alternaria brassicicola), Sclerotinia stem rot (Sclerotinia sclerotiorum), blackleg (Leptosphaeria maculans), and Fusarium wilt (Fusarium oxysporum) (Morita et al. 2003). To prevent the misdiagnoses, the proper diagnosis targets on the root systems, which is the name sake of the disease itself “clubroot disease”. By digging up the roots and searching for the galls, clubs, tumours, swelling tissues or distorted root shapes on the infected plants. Galls appear as spindle like shredded, knobby, or club shaped swelling varied by the growth stages and resistant-susceptible status of the hosts. Oftentimes, some of the galls go decayed in the matured plants but leave the roots with brown peaty appearance rather than a healthy white colour. When cut out, the interior texture of a clubroot gall is spongy or marbled.

1.1.7 Histological assessment and microscopic examination

All development stages of the pathogen *P. brassicae* can be microscopically observed with the aid of several staining methods. Stages such as primary

zoospores and secondary zoospores are similar when examined under the microspore. These stages can be differentiated by locations on root hairs or main roots to decide whether it is primary or secondary zoospores respectively. Resting spores are easily observable by clusters densely located in the roots. The secondary infection has several distinctive features when stained. Plasmodium, zoospores migration, cellular activities such as hyperplasia (cell division) and hypertrophy (cell expansion), etc… can be observed clearly (Dixon 1991). For instance, lipid droplets can be visible with large size within the clusters of zoosporangia because this lipid storage serves as the energy reservoir for the pathogen (Bi et al. 2016). Apart from the observable invasive plasmodium intracellularly, cytological changes can also be visible under the microscope. The affected host cells are enlarged and become circular in shape instead of square-like. Cell walls gets thinner and cellular organelles are inconspicuous in the heavily infected cells but their nucleus are dense and larger due to hypertrophic activity induced by the invasive plasmodia (Dixon 1991).

1.1.8 Genome of *P. brassicae*

Genome of *P. brassicae* has been sequenced by several groups (Siemens et al. 2009, Rolfe et al. 2016, Schwelm et al. 2016, Daval et al. 2019). The first genome sequenced was the European SSI e3 isolate published in 2015 (e3_2015), then came additional re-sequencing of five Canadian, and one Chinese isolates and recently 48 genomes were assembled and deposited in NCBI genbank (Schwelm and Ludwig-Müller 2021). As an intracellular obligate parasite, *P. brassicae* genome is quite compacted and lack of significant intergenic spaces. However, spliceosomal introns were present in genes as Bulma et al. (2007) pointed out in their research of 24 intron-carried genes detected in the *P. brassicae* genome. Genes are prioritized for obligate biotrophic parasitism such as hormonal regulation of the host growth, other genes responsible for the essential nutrients such as biosysthesis of thiamine otherwise are lost since the

nutrients supplied by the host, rendering *P. brassicae* a complete dependency of an biotrophic parasite (Rolfe et al. 2016). Highlighted genes during the life cycle of *P. brassicae* can be counted as Chitin synthesis essential for spore germination; basal and lipid production for infection process; cytokinin biosynthesis, auxin homeostasis, salicylic acid, and jasmonic acid metabolism for hormonal modification in host.

1.1.9 The challenges of clubroot study

There are several challenges for approaching the clubroot disease for both aspects from the hosts to pathogens. For the host plants, studies on Brassica crops for breeding and selection of resistant lines are also a remedy for crop industry. For the pathogen *P. brassicae*, the greatest challenge is its nature as an obligate parasite which is non-axenic, in other words *P. brassicae* cannot be cultured independently (Schwelm et al. 2016). The clubroot experiments must involve in "*in vivo"* (cultured in living organisms) activities which are laborious. For instance, single-spore isolate is always preferable in *P. brassicae* molecular research, but the making-single-spore process is time-consuming and difficult for clubroot experts due to their microscopic size.

Due to the complex life cycle with asexual and sexual reproduction, the wild *P. brassicae* populations are comprised of variable isolates, races and pathotypes even in one gall. These definitions can generally be understood as pathotype is a group of organisms that have the same pathogenicity on a specified host (Hwang et al. 2012); isolates or strains are individual populations and races includes multiple isolates. To dissect and analyse them, it requires more than techniques, skills and related experiences, particularly for the molecular investigation. For instance, the *P. brassicae* research of two groups from Korea (Laila et al. 2017) and Japan (Niwa et al. 2011) reported the new geographical populations of *P. brassicae* based on the ribosomal DNA regions but those findings were repudiated by Schwelm et al. (2016, 2017) because the genetic makeup was

misidentified with other soil borne protists contaminated in the samples. Indeed, the technology might be advanced yet the relevant information about *P. brassicae* and its protist relatives remain insufficient. In fact, *P. brassicae* is among the first genome sequenced within the Rhizaria group, until now only other plant parasitic members such as *Spongospora subterranea* and *Polymyxa* species whose genomes were fully sequenced (Schwelm et al. 2017). The other less economically important members such as the free-living protists in soil, water or sea whose genetic information is still limited. Therefore, clubroot experts must be cautious when it comes to approach *P. brassicae* molecular research.

Another the greatest challenges for studying *P. brassicae* is the classification of pathotypes. Up to date, there is no unification for the determination of clubroot pathotype. Since each nation has its own interest on varieties of Brassica crops grown, geographical conditions and history of clubroot occurrence, thus several differentials sets have been proposed. These include Williams (1966), Some et al. (1996), European clubroot differential (ECD) set (Buczacki et al. 1975) were among the first one. Then, they have been adapted to be suitable for each country. For instance, Canadian clubroot differential set (CCD) (Strelkov et al. 2018) and Sinitic clubroot differential (SCD) were developed to meet the standard of pathotype characterization on their own countries. Howevermuch effort put in the physiologic specialization, the clubroot experts are still scratching their head on this topic because the deep down problem involved with molecular linkages among the pathotype chaos is another level. As mentioned above, due to the nature of being an intracellular parasite, purification of the *P. brassicae* DNA is of a hard task. For the accuracy, it requires multiple factors such as single-spore isolate and specific primers to prevent the contamination of host DNA or other soilborne microbes. In addition to the chaos of the pathogenicity tests, the efforts made on finding the molecular linkages correlated to those pathotypes seem unsolved. There were reports on molecular makers that are able to distinguish the

races and pathotypes, i.e. Williams pathotypes P11, P9, P7 and P4 (Zheng et al. 2018); P5 (Feng et al. 2016) predominant in China; P5X newly emerged in Canada (Zhou et al. 2018), etc... However, these markers seem restricted to geographical and local conditions and few are unverifiable due to the incomplete genomic data (Schwelm and Ludwig-Müller 2021). Overall, the clubroot pathologists have been yet comprehended the dynamics or related molecular mechanisms driven behind the pathogenesis of these pathotypes. A set of standard markers is in need to describe pathotypes and map their distribution, that renders for several beneficiaries such as clubroot experts and Brasscia breeders to study the dynamics of pathogeneticity (Siemens et al. 2009).

Last but not least, effective control of clubroot disease has been proven as a difficult task. In soil, the thick-walled resting spores stays decades and easily transported by agricultural activities, animal, infested soil and water… The management of clubroot is thus based on integrated management combined several strategies including soil preparation, crop rotation, …to avoid the severity and widespread of the disease because chemical controls have been proven either minor effective or harmful to the environments (Struck et al. 2022). One of the most effective strategies of clubroot control is the deployment of the resistant breeds based on the CR genes (Clubroot Resistance) of Brassicaea family (Jiang et al. 2022). Nevertheless, the battle between hosts and pathogens is a never-ending game because the new pathotypes or races are always arisen when the resistant lines are grown. The *P. brassicae* is famous for pathogenicity alteration due to its genetic diversity that contains variations that are able to supress the resistance sources (Diederichsen et al. 2009). This factor renders the war against the protist *P. brassicae* endless.

1.2.1 The cruciferous weed *Cardamine occulta*

The *P. brassicae* populations investigated in this research were extracted from the host weed *Cardamine occulta*. As stated above, our approach was

different from other mainstream studies, mostly on Brassica crops. In general, the taxonomic status of this cruciferous weed has been also currently certified as a new species, since this weed has a long history of misidentification with other members of the *Cardamine* spp. group.

In this research, we were not yet to document the morphological and anatomical characters of this weed but pictures can be found at these websites Plants of the World Online (https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:280533-1) or Flora of New Zealand (https://www.nzflora.info/factsheet/taxon/Cardamine-occulta.html) with full descriptions from Heenan (2017). We briefly describe the *C. occulta* taxonomic status, related genomic information and most importantly its life cycle coevolving with rice cultivation. These elements are foundations for the hypothesis that revolves around the evolutionary relationship between the weed *C. occulta* and the protist *P. brassicae*. In this host-pathogen relationship, the weed host acts more than a competent reservoir to *P. brassicae.* The host harbors great level of genetic variations and genomic diversity rooted from polyploidy genome (Liu et al. 2010), this enriched genetic materials are driving forces to diversify the genetic makeup of *P. brassicae* at population level via the host-pathogen interaction and evolution.

1.2.2 Nomenclature

Cardamine occulta has had a long history of misidentification by its close relatives *C. flexuosa*, *C. hirsuta*, *C. kokaiensis* due to the morphological similarities and growing habitats. *C. occulta* was traditionally considered as Eastern Asia population (Asian *C. flexuosa*) or a subpopulation of *C. flexuosa* (Marhold et al. 2016) due to the fact that *C. flexuosa* is a very common weed grown in Europe. Currently, the name *C. occulta* is not widely accepted (e.g. the World Checklist of Seed Plants (Govaerts, 1999) and still preferred by the *C. flexuosa* in several publications.

The morphological differences between *C. occulta* and other common *Cardamine spp.* have yet fully documented. *C. occulta* is claimed to have distinguishable characters from other typical *C. flexuosa*, *C. scutata*, *C. hirsute,* etc …. For instance, the *C. occulta* weed does not form a distinct rosette nor have scattered hairs on the upper leaf surface as oppsed to that *C. flexuosa* does. The *C. occulta* has hairy flower buds and auxiliary young shoots which are absent in other Cardamine species (Hruševar et al. 2021). Nevertheless, accurate identification to separate these closely-related weeds is never an easy task for a layman or even clubroot experts, especially with the coexistence situation that these weeds regularly share similar habitats, i.e. *C. scutata* is found in the water-clogged rice fields which *C. occulta* also flourishes in (Marhold et al. 2016).

Most of the descriptions come from publications or guidelines that described *C. occulta* as an invasive species (Ardenghi and Mossini 2014, Mansanet-Salvador et al. 2015, Heenan 2017) so as to monitor its status. However, empirical evidences of genetic makeup reported by Lihová et al. (2006a) and Marhold et al. (2016) confirmed *C. occulta* is a new species with a distinct octoploid genome.

1.2.3 The polyploidy genome

The *C. occulta* weed possesses an allopolyploid genome (Lihová et al. 2006, Slenker et al. 2018). The term “allopolypoidy” is to describe the origin of the gemone derived from two different species. In case of *C. occulta*, Mandáková et al. 2014 proposed it might originate from the diploids *C. kokaiensis* and *C. scutata* ($2n = 2x = 16$) based on the comparative genome painting method. The genome is an octoploidy (2n=8x=64) whose sets of chromosomes are double of the closest relatives allotetraploid *C. scutata* and *C. flexuosa* (2n=4x=32, two sets of homologous chromosomes). The duplication of the entire genome is well known as polyploidization evolution that occurs commonly in the *Brassicaea* family across over domesticated crops as well as wild crucifers (Kagale et al.

2014). Among the *Cardamine* species, the polyploidy events are main forces driving the speciation in *Cardamine* (Marhold et al. 2016). In *C. scutata* whose tetraploid genome comes from the parental diploids *C. amara* and *C. parviflora* that are specialized to different conditions. *C. amara* is known as a waterlogged weed but *C. parviflora* prefers in drier conditions (Shimizu-Inatsugi et al., 2017). The combination of parental genomes provides the progeny *C. scutata* the evolutionary advantages of being adaptable to broader range of water-fluctuated habitats. Likewise, the genomic duplication occured in *C. occulta* weed is probably a hybridized progeny of *C. scutata* and *C. kokaensis*, thus allows *C. occulta* expand efficiently on a greater spectrum of ecological niches, particularly for anthropological conditions, e.g. rice-paddy fields. The *C. occulta* weed is well-established in the man-made habitats proves the most effective measure is to explore and widespread to newly ecological niches since agricultural activities can introduce them to any continents (Lihová et al. 2006).

1.2.4 The life cycle of *C. occulta* in rice paddy fields

The life cycle of *C. occulta* is considerably dependent on rice cultivation, summarized in Fig.3. As elaborated in the section of *C. occulta* genome, the evolutionary advantages inherited from parental genomes strengthen the capacity of being adapted to the human-induced conditions such as rice fields, urban areas, construction sites, etc…. (Lihová et al. 2006). The genomic data presented in Markhold et al. (2016) suggested that *C. occulta* genome emerged later than that of *C. flexuosa* and the origin of *C. occulta* was perhaps associated with history of rice cultivation.

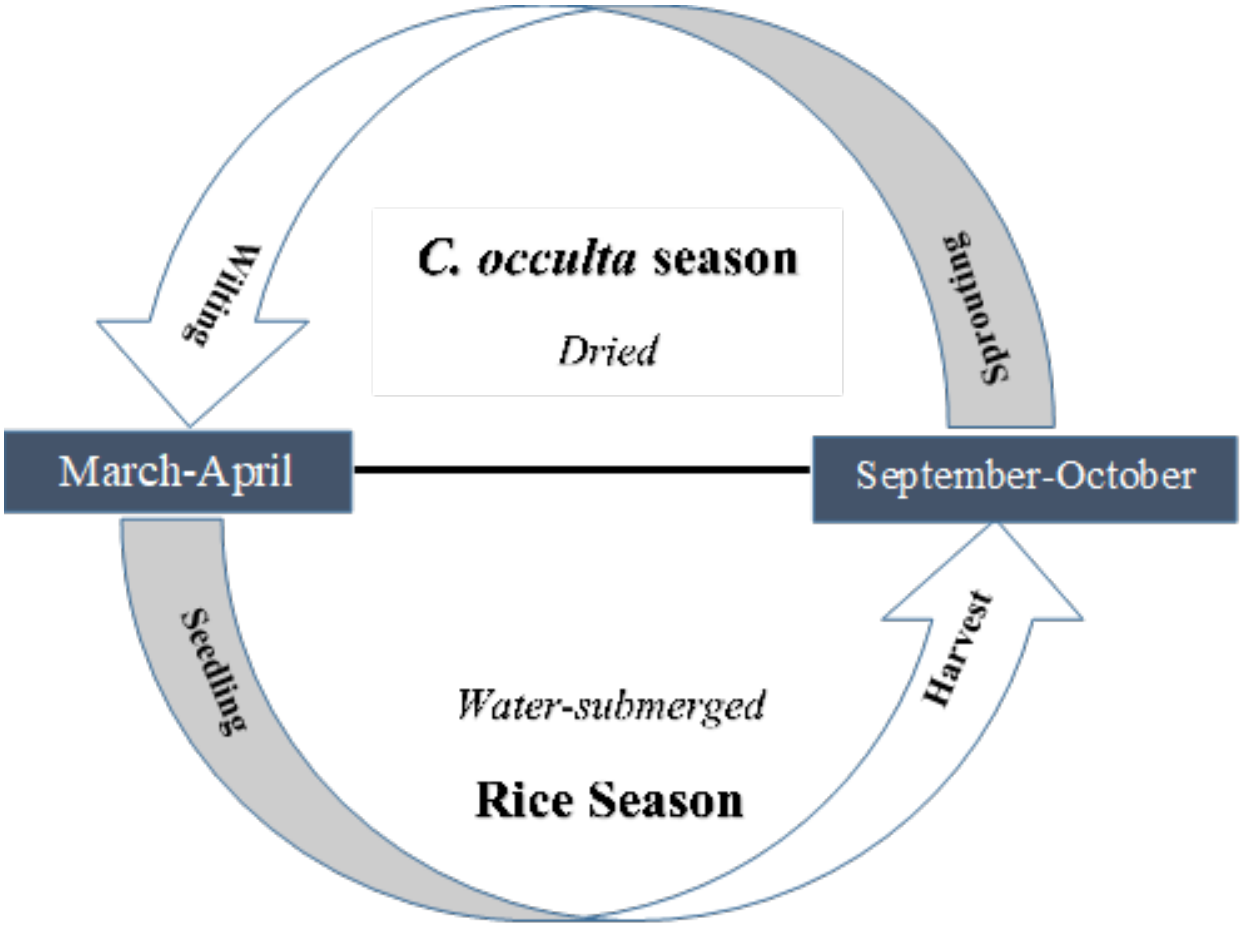


Figure 3. Summarized life cycle of *C. occulta*. The graphic demonstrates the period times when the the weeds start sprouting simultaneously during the harverst time of the rice (September-October); and the life of the weed ends where the new rice season begins (March-April). Soil conditions are indicated Dried and Water-submergred as hydrologic fluctuation.

The usual life cycle of *Cardamine occulta* here in Japan commences at the time of autumn (September-October) when rice is ready to harvest. Sprouting of the *C. occulta* weeds can be spotted early at the edges of the fields where exist the irrigated systems. At the post-harvest time, fields are abandoned then the

weed seedlings observed are increased in numbers. Usually if the post-harvest rice fields are drained and cleared of stubbles and debris, a greater number of *C. occulta* can be easily found. Seedlings and mature plants are abundant after November and the numbers of plant in population peaks at the period from December to February (Fig.4) depending on the natural and climatic conditions. For instance, as observed the annually-operated fields had greater number of *C. occulta* weeds counted than those weeds grown in the abandoned ones perhaps due to the fact that rice fields are taken in good care or highly-precipitated and warmer weather also resulted in higher numbers of the *C. occulta* population than the drier and colder condition. After the peak time of population, plants started to senesce and wilt during the spring time (March-April). The spring time signals the beginning of rice season. Subsequently, the rice fields again are irrigated with water, thus the *C. occulta* seeds stay dormant in the submerged condition awaiting for the end of rice season. When the rice fields are emptied and drained again, the life cycle of *C. oculta* weed restarts. The seeds obtain the advantageous characters that allow them alternatively stay dormant or germinate on occasionally submerged conditions. These characters enhance the survival ability and invasiveness of this weed (Yatsu et al. 2003).

Figure 4. Images of *C. occulta* infested in the rice fields at the highest number of populations (Photos taken in December and January). The was all blooming with white flowers as well as seed pods

1.3 OBJECTIVES

With the advent of molecular techniques and knowledge, clubroot disease has been intensively studied. However, considered the timeline of *P. brassicae* genome sequencing, this disease is still a new player in the game. Controversies about the effectors, pathogenesis, molecular links among the pathotypes, sexual reproduction etc… are ongoing heated-bated. Indeed, there still several pieces unsolved in the clubroot puzzle that awaits to be assembled. Our research is aimed to approach another piece on this clubroot puzzle, we focus primarily on the cruciferous weed *C. occulta* instead of mainstream research on Brassica crops. These weeds might seem trivial to the clubroot streamline because they do not bring any benefits to the economic aspects. However, the relationship between the *P. brassicae* populations intertwining with these weeds might divulge the story of host-pathogen coevolution which might involve in pathogenesis, and population-level diversity. These elements are likely to contribute significantly to the dynamics of the holistic clubroot research. To elucidate about the novel cubroot pathogen on *C. occulta* weeds, two aspects were examined thoroughly

1. *The primary aspect is phenotyping the pathological characters and distinctive traits (morphological, histopathological, and anatomical) that feature on the clubroot disease of C. occulta weeds, how these characters are different from those of on Brassica crops.*

2. *The secondary aspect is genotyping the P. brassicae populations from the weeds based on the certain genes such as ITS, Actin and intronic regions of the rDNA. These empirical evidences prove that the genetic materials of those P. brassicae on weeds is far distinct from the cosmopolitan strains.*

CHAPTER 2

MORPHOLOGICAL AND HISTOPATHOLOGICAL ASSESSMENT ON THE CLUBROOT DISEASE OF *C. OCCULTA*

2.1 INTRODUCTION

The stereotypical characters of clubroot disease on Brassica crops were describe in Chapter 1 with the above-ground symptoms and more importantly the galls formed at the root systems. In Brassica crops, severely clubroot affected plants show both above and underground symptoms which are observable by discolored and thrifty plant with uprooted large galls. For clubroot disease in the wild weeds, generally speaking there were yet reports to describe unequivocal symptoms since either the wild weeds are economically trivial or resistant to the pathogen *P. brassicae*. Symptoms on the wild weeds often go unnoticed because most of cases are symptomless or mild symptoms which assumed as other diseases harbored by the weeds. Oftentimes, molecular tests by specific primers should be deployed to affirm as clubroot disease. Ren et at. (2016) demonstrated the host range of *P. brassicae* including cruciferous and non-cruciferous weeds, pot trials and field surveys proved that all the members of cruciferous weeds can be hosts for *P. brassicae*. However, the symptoms for the weeds were minor with mini-galls, even in the case of *Capsella bursa-pastoris* a common weed in China was the most susceptible among the weeds, incidences were scarce and random (Kim et al. 2011, Ren et al. 2014). The first clubroot reports on *C. occulta* (under the name *C. flexuosa*) was conducted by our group with descriptions on distribution, the rate of infection and symptomology with only the appearance yet anatomical or histopathological demonstration. In these reports (Tanaka et al. 1993, 2006), clubroot occurrence was detected all over Japan (43 provinces), far reaching to the most desolated islands except for Hokkaido, the northernmost of

Japan's main islands. The reason is perhaps the climate in Hokkaido is harsh and unsuitable for growing rice, thus *C. occulta* whose life cycle relied heavily on rice is not present accordingly. Clubroot on *C. occulta* in Japan is more conspicuous and widespread compared to other research on clubroot of cruciferous weeds. The frequencies and incidences of clubroot on the *C. occulta* weeds were relatively high, in some cases a half of locations within surveyed area observed with clubroot and up to 85% infected weeds per sampling, respectively (Tanaka et al. 2006). The overabundance combined with the highly-infected rate of this wild weed thus hypothesizes it might play an active host role more than an alternative reservoir for the protist *P. brassicae*.

In this research, we repeated briefly the morphological description but supplemented the histopathological demonstrations on the secondary infection stage in a sense of comparative analysis with typical symptoms on Brassica crops. The clubroot symptoms on the *C. occulta* weed were in fact more distinctive than other cruciferous weeds', and most importantly different from the Brassica crops. Briefly, the above-ground symptoms of the *C. occulta* clubroot also goes asymptomic in despite of the heavily infected weeds. For underground symptoms, galls instead of forming in the root systems are detected on the stems (stalks) of the weeds, and lastly when examined under the microscope the plasmodia were detected in the stele, vascular bundles or even in the petioles.

2.2 MATERIALS AND METHODS

2.2.1 Field surveys on clubroot disease of *C. occulta*

For pathological observation and morphological, histological description of clubroot on *C. occulta*. A survey was conducted on rice fields within Yamaguchi city during 2018-2020. Approximately 50-100 plants were uprooted and

examined for their clubroot gall formation at random fields. Plants at various stages of development were examined during their life cycle from September to April. Following the methods of Tanaka et al. (2006), infected plants were investigated according to rates of visible symptoms and also through microscopic observations. Few samples collected from the fields were also PCR-checked for clubroot disease but data not shown in this article.

2.2.2 Clubroot samples collected

For genetic investigation, we used the clubroot samples collected from diseased plants between the period 1994-2014 from multiple locations across Japan (Table. 1). The collected clubroot galls were washed thoroughly and stored at -40°C.

2.2.3 Microscopic observation

Plant tissues (hypocotyl, stem and root) were firstly diagnosed by pathological characteristics including spheroid, spindled large galls and fixated with 2% glutaraldehyde, and then subjected to cryosectioning. Sections obtained were primarily observed under a BH2 light microscope (Olympus, Tokyo, Japan) to confirm the presence of *P. brassicae*. Some sections were stained with 1% acetocarmine before the light microscopy. For fluorescent microscopy, the plant tissues were sectioned and stained with 10 μg/mL Nile red and observed under a fluorescence microscope (BIOREVO BZ-9000, Keyence, Osaka, Japan) to illustrate the infection stages of *P. brassicae*.

2.2.4 Preparation of resin-embedded specimens and sections

Plant tissues were fixed with 2.5% glutaraldehyde in 0.05M sodium phosphate buffer (pH 7.0) for 2h at room temperature and rinsed with the buffer

for 2h. These fixed tissues were dehydrated in an increasing ethanol series (50%–100%) and embedded in Spurr's resin (Spurr 1969) after two treatments with 100% propylene oxide. Embedded tissues were cross-sectioned (about 1.2µm in thickness) with a glass knife on a PorterBlum MT-2 ultramicrotome (Ivan Sorval, Norwalk, CT, USA). These sections were stained with 1% toluidine blue for observation by light microscopy.

2.3 RESULTS

2.3.1 Characterization of Clubroot Symptoms on *C. occulta*

Clubroot symptoms on *C. occulta* vary greatly depending upon the months in which the examination was undertaken and the growth phases of the host weed. The above-ground symptoms of the diseased plants were mild to virtually asymptomatic despite having severe infection symptoms in the root systems, therefore diseased plants became hardly diagnosable in their natural habitat.

Certain *C. occulta* populations showed minor morphological differences such as early flowering including bolting and stalking and less pronounced rosette structure. The underground symptoms appeared less developed in several root systems (lateral roots and main roots) and numerous massively deformed gall shapes e.g., club-, spindle- and spheroid-shaped root tissues (Fig. 5a, c), the occurrence of necrotic tissues were macroscopically visible as brown spots on root surfaces (Fig. 5a) or on stems and hypocotyls (Fig. 5c, b, d).

In September and late October when the weeds are in germination, only 10-20% of the young plants were found infected. In the following months, both the infection rate and the severity of clubroot disease increased proportionally to the surge of the weed population. The percentage of infected plants peaked between December to February, ranging from 60 to over 80%, simultaneously the time of *C. occulta* population flourished with blooming and seeding. The weeds dried out

over the summer months and concurrently the clubroot disease declined with only 20-30% of plants diagnosed as being diseased. Mini-galls and small nodules, regarded as benign symptoms, have regularly been observed in the weed sprout and wilting period. Large tumors were identified during the peak time, particularly in stem and upper parts of the weeds, which were conspicuous during this period. Galls were frequently detected in hypocotyls, adventitious buds, and even upper parts such as petioles at several stages of the host weeds (young, pre- and mature), multiple stem galls could easily be noticed above ground with bulb-shaped structures (Fig. 5c, d). In young plants, stem galls are formed cylindrically and are connected directly to the root systems, wherein later in the weed development, the adventitious roots, shoots and stolons grow out from the infected zones (Fig. 5b, d).

Figure 5. Clubroot disease symptoms in root systems of *C. occulta*. (a) Young plants sprouting in October with spindled main root. (b),(c) Premature plants in November with spindled main root and spheroid tap roots. (d),(e) asymptomatically above-ground plants with (d) developed root systems with tiny galls and brown spots at root tips and (e) less developed root systems filled with large nodules. (f),(g) putatively susceptible *C. occulta* population with brown peaty appearance roots and minor above-ground symptoms of early flowering and stalking. Bars represent 1cm for all panels.

Figure 6. Clubroot disease in stems and hypocotyls of *C. occulta* (a) Severely infected plant with large tumors and decayed tissues. (b) Young shoot sprouting from infected part (hypocotyls presumably). (c), (d) bulb-shape stem galls, hollow core with decayed out-layer (e),(f),(g),(h) Clubroot on young and mature plants found between October and December with cylindrical shaped from main root to stem bases. Bars represent 1cm for (a)(b)(c)(d) and (g)

2.3.2 Histological Analysis under Microscopic investigations

Primary infection in root hairs could not be conclusively determined since the root hairs harbor plethora of microbes other than *P. brassicae* in natural habitats, although greenhouse experiments (data not shown) indicated similar process to *P. brassicae* of the crop group.

Secondary infection stages were observed in roots and disparate parts of the host (Fig. 6). At the emergence as well as terminal stage of the *C. occulta* weeds, spore-forming plasmodia and resting spores were frequently observed in root tissues ((Fig. 6a, d). At the peaked time, the existence of various vegetative plasmodial stages in the growth process of *P. brassicae* could be detected at different locations of the specimens examined young shoots (Fig. 7b), stems (Fig. 7a, c), and pith (Fig. 7d). Generally, the stages are similar to those of the *Brassica* crops (Riascos et al. 2010, Schuller and Ludwig-Müller 2016), i.e., stimulating the hyperresponsive reactions from host cells, lipids and starch grains accumulation.

The majority of identified cases for the plasmodial invasion and colonization in the *C. occulta* tissues (roots, stems and other parts) appeared mildly damaged as opposed to the invasion in the cruciferous crops which usually causes more distortions to the plant tissues (e.g., decayed roots tips, rotten tissues). Nile red-stained lipid components were extensively visible under a florescent microscope by the distribution of lipid contents throughout the cytosol. Despite severely infected conditions in stems and hypocotyls, young and mature plasmodia were only seen in meristem, occasionally in phloem and epidermal cells (Fig. 8b). Although the general stem anatomy remained intact, infected cells displayed the cytological features of cellular hypertrophy and hyperplasia (Fig. 8a, c). In meristem cells, lipid droplets stained by Nile red of plasmodia were observed under fluorescence microscopy (Fig. 8d). However, resting spores and sporulating plasmodia were rarely noticed in pith region and vascular systems.

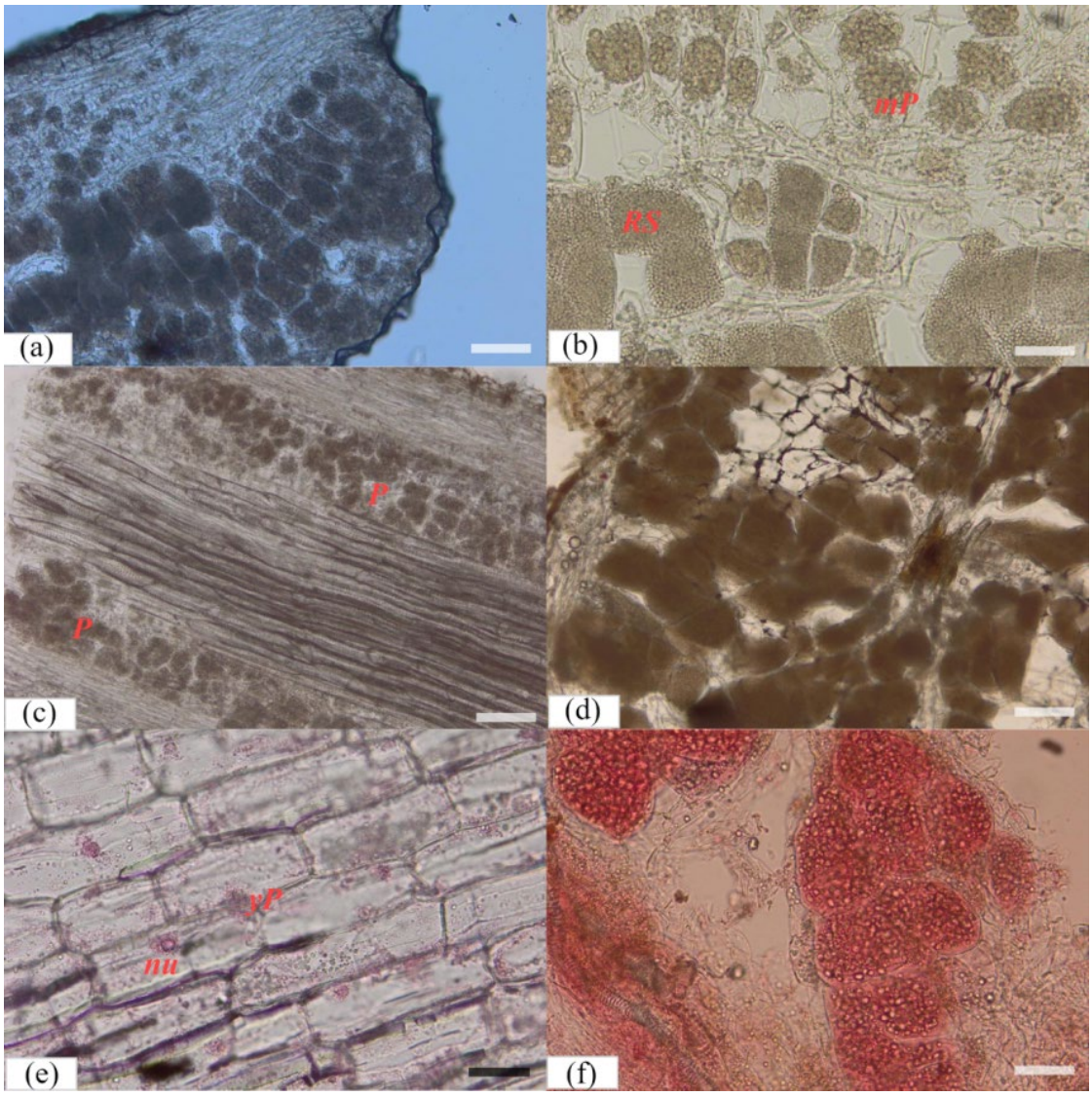


Figure 6. Conventional microscopic observation on infected root tissues of *C. occulta*

(a) Sporulating plasmodia in a longitudinal-section spheroid gall

(b) Sprulating plasmodia containing resting spores (RS, resting spores) and mature plasmodia (mP) .

(c) Longitudinal section of root showing plasmodial invasion (P) delimited from the secondary cortex.

(d) Cross section of a small gall occupied by resting spores.

(e) Hypersensitive response of secondary cortex cells with enlarged nuclei stained with Nile red method. nu, nucleus; yP, young plasmodia

(f) Cluster of mature plasmodia stained with acetocarmine.

Bars represent 20 μm in panels (a), 10 μm in panels (b)(c)(d)(e) and (f)

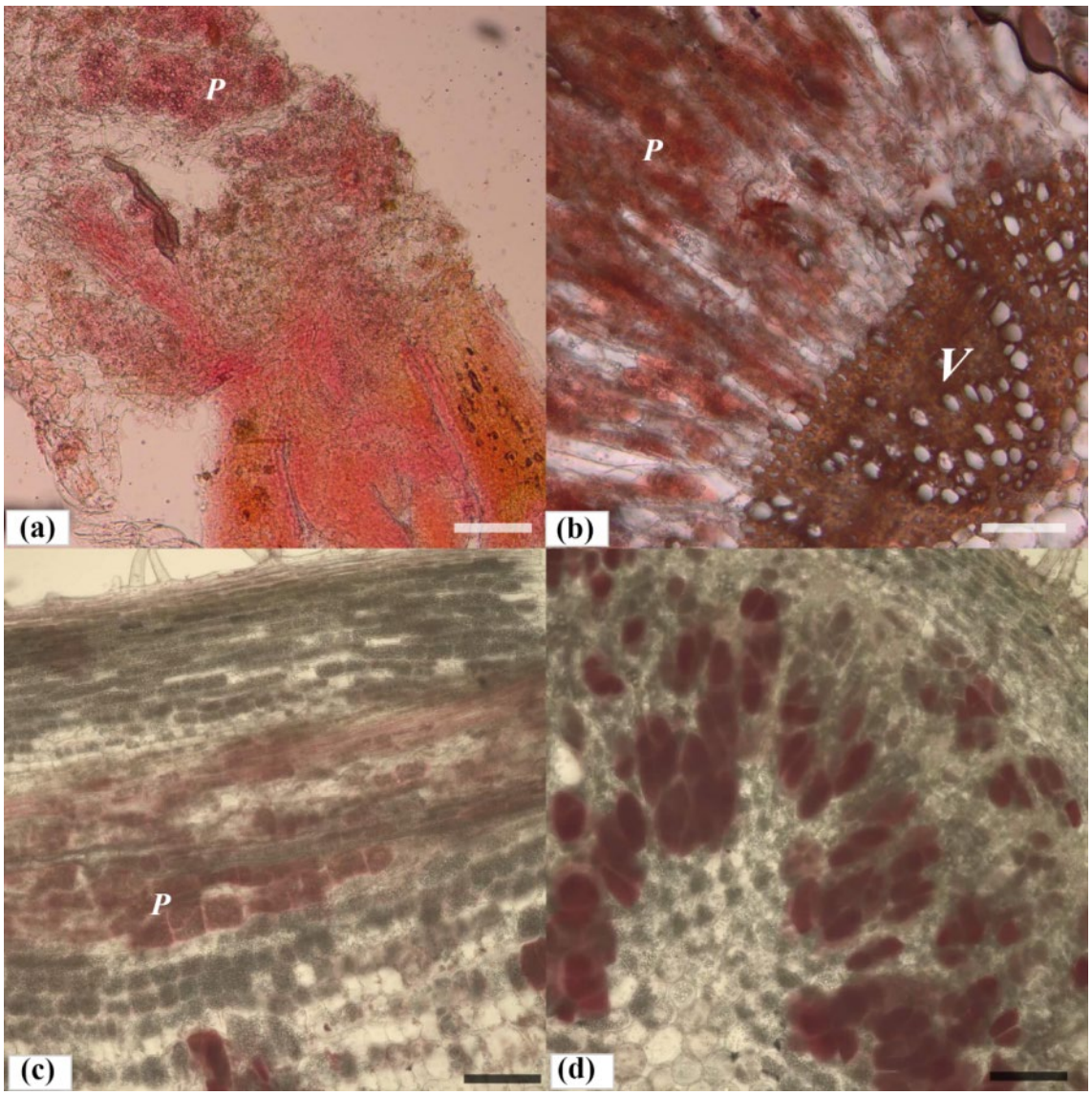


Figure 7. Conventional microscopic observation on infected stem tissues of *C. occulta*. (a) Plasmodial invasion in young shoot. (b) Cross-section and (c) longitudinal section of stem infected with plasmodia in parenchyma cells of the secondary cortex. (d) Plasmodia found in the pith. All samples were stained with Nile Red. The unstained (dark spots) accumulated in cells were identified as starch grains. P, plasmodia; V, vascular bundle. Bars represent 100 μm in panels (a) and 20 μm in panels (b)(c) and (d)

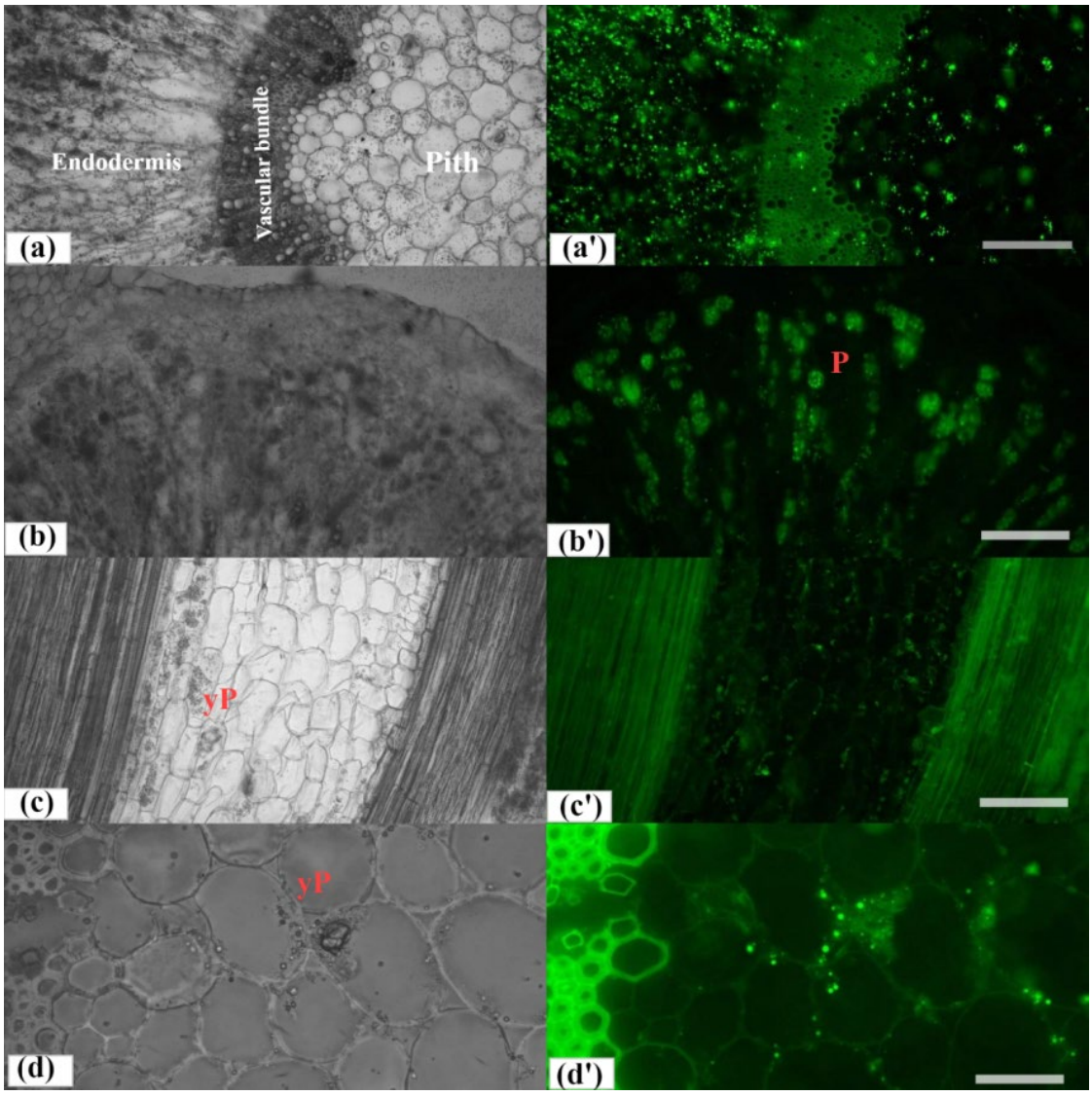


Figure 8. Comparative paired images of *P. brassicae* infecting *C. occulta* stems visualized bytly mature secondary plasmodia, sporulating secondary plasmodia and resting spores can be identified. Plasmodial structures display a yellow-green autofluorescence

(a,a') Secondary plasmodia presence in hyperplasia and hypertrophied parenchyma cells across stem structures from endodermis to pith.

(b,b') Hypertrophied (cell enlargement) parenchyma cells in medullar rays (P, plasmodia)

(c,c') Longitudinal section and (d,d') Cross section of young plasmodia invasion in pith, notice (d,d') the cytoplasmic streaming of myxamoebae-like structure to the adjacentcells wall (yP, young Plasmodia)

Bars represent 100 μm in panels (a)–(d), and 200 μm in panels (e) and (f)

2.4 DISCUSSION

Reports have shown that *P. brassicae* are able to invade a wide range of hosts even with non-Brassica plants but the macroscopic symptoms were rarely reported in nature (Ren et al. 2016). For *C. occulta*, as a wild cruciferous weed, it acquires a few visibly mild symptoms, e.g., galls in root and stem, and under microscopic level, e.g., various plasmodium stages and resting spores. Our theory is this could be an indicative of an ecological adaptation of the pathogens to the new natural habitats. The *C. occulta* weed previously reported as *C. flexuosa* is a native East Asian weed whose genetic and morphological taxonomy is still ongoing research and discussion (Marhold et al. 2016). In Japan, the *C. occulta* weeds are considered as a part of weed community established in the rice cultivation process. They thrive and proliferate in the well-irrigated paddy fields. Their life cycle seems to be coevolutionary with rice-growing environments in contrast to those conditions facilitating the Brassica crops.

This is the first study that documented the morphological and histopathological description on the *C. occulta* weeds. Based on the symptoms observed, it might appear that the pathogens were not that harmful to the host weeds. Those characters such as swollen hypocotyls, spheroid galls… are similar to those occurred in the resistant hosts described by Liu et al. (2020). Indeed, it could be an assumption that the wild weeds whose polyploidy genomes are enriched with a great number of genetic variation including the resistant genes, thus the pathogens are necessarily compromised to reach to the point of mutualistic relationship. As a result, we only observed coexistence between host-pathogen. Even in case of the deadly virulent pathogen existing, natural selection takes place as eliminating the susceptible individuals from the weed populations. Consequently, the host-pathogen interactions gradually select the healthy weeds (resistant or less susceptible) and avirulent pathogens cohabitate in peace.

CHAPTER 3

GENETIC INVESTIGATIONS ON ITS REGIONS AND ACTIN GENES OF *P. BRASSICAE* REVEALING THE NOVEL POPULATIONS

3.1 INTRODUCTION

As demonstrated in previous chapters, the clubroot occurrence in *C. occulta* have been recorded unusually high incidences all over Japan. In addition to the incidences and frequencies, overwhelming evidences of the morphological and histopathological analyses also raised the questions of this weed-pathogen relationship at molecular level. In this chapter, we venture to examine the genetic materials of the *P. brassicae* infecting *C. occulta* in order to search for the differences separating this special group from the cosmopolitan *P. brassicae.*

In this perspective, *actin* genes and ITS (Internal Transcribed Spacer 1 and 2) regions have been selected as key markers for genetic investigation of *P. brassicae* populations. Actin genes were selected based on previous research (Archibald et al. 2005). For the protists, actin genes play vital roles in several essential functions of life stages such as cytoskeletal structure, maintenance of cell shape, cell motility, cell division, endocytosis and intracellular transport (Yi et al., 2016). In *P. brassicae*, Actins also participate in the pathogenesis process as in the function of mobility of the flagellated zoospores swimming and attacking the root systems. Fundementally, Actin genes are the household genes whose mutation rates are less likely to occur, thus any minor changes e.g. silent mutations signal the shift of genetic at the population level. Archibald et al. 2005 reported the Japanese population (one of the strain sent from our laboratory) had mutations different from other global strains. This is the reason that Actins have been selected to aim for the genetic differences between the *P. brassicae* on *C. occulta* (Cardamine group) and cosmopolitan strains on Brassica crops (Brassica group).

ITS regions are the intergenic or non-transcribed DNA sequences located between the large subunit (LSU) and small subunit (SSU) of ribosomal RNA. More specifically, for *P. brassicae* and other eukaryotes ITS1 situated between 18S and 5.8S and ITS2 between 5.8S and 28S of the rRNA genes. The functions of the ITS regions is still controversial but they were considered as RNA precusors excision and regulations (Prahl et al. 2021). Nevertheless, ITS regions have been broadly employed in molecular phylogeny due to their high degree of genetic variations and mutations for elucidating the phylogenetic relationships between closely related species. Laila et al. (2017) utilized ribosomal DNA sequence polymorphisms including ITS regions to distinguish the *P. brassicae* geographical populations revealing the molecular characterization of Korean *P. brassicae* populations different from the cosmopolitan strains. However, the results from Laila et al. (2017) was still controversial since the cosmopolitan *P. brassicae* has been proved to be conserved in the ITS regions. A further step to clarify the distinction of ITS2 region by analyzing its secondary structure. This method has been applied to prove the closedly related species with only a few mutations detected in the ITS2 regions. These mutations when formed (by base-pair interactions between canonical base-pairs (AU, GC), non-canonical stable (GU), unstable (AC)) might create extra structure or even change the canonical structure (four helices), revealing additional phylogenetic information usually not found in the primary sequence, thus including this information can significantly improve phylogenetic estimates (Keller et al. 2010). In this investigation, we used the ITS2 database website created by Jörg et al. (2006) to predict the secondary structures of the Japanese strains. This advantageous tool allows for predicting the formation of secondary structures of ITS2 used by cellular activities in the properly folding process of the pre-rRNA into Ribosomes. Therefore, these proposedly formed structures display several dissimilar

characteristics within monophyletic clade, thus indicating higher taxonomic richness and population divergences.

3.2 MATERIALS AND METHODS

3.2.1. DNA Extraction

The clubroot samples were washed thoroughly by tap water and then distilled water. Resting spores of *P. brassicae* were isolated as described previously by Tanaka et al. (1990) by repeated centrifugation of a spore suspension until the supernatant became completely transparent. DNA from resting spores of each *P. brassicae* population was isolated as described by Ito et al. (1997).

3.2.2 Primer designation, Cloning and Sequencing

All primers were designed by Primer3 software (Available online: http://primer3.ut.ee/). *Actin* primers were designed based on DNA sequence AM411664, ITS primers based on KX011115 in GenBank database (Table. 1). General PCR reaction conditions were in Table 1. PCR amplification were carried in a 50µL reaction mixture (Takara Taq) containing 1 µL gDNA, 2 µL each primer, 1 µL Takara Polymerase, 10 µL Takara Buffer, 4 µL dNTPs and 32 µL distilled water. PCR products were cloned into the pGEM-T Easy cloning vector (Promega, MD, USA). Transformation of competent *E. coli* JM109 competent cells Sequencing of the cloned DNA was performed with the primers SP6 and T7, the binding sites of which are upstream and downstream from the insertion site in the vector. Products were sequenced by ABI Prism 3700 (Applied Biosystems, CA, USA).

Table 1. List of primers used in this research with PCR conditions, products size and targeted regions.

Markers	Forward (Fw) and Reverse (Rv) Primers	Products size (bp)	PCR condition
Actin	Fw CGCAGATCATGTTCGAGACG Rv CGGACGGGTAATACTGCTCA	787	96 C 3 min, 96 C 30s, 56 C 45 s 72 C 1 min, 32 cycles, 72°C 10 min
ITS (Partial 18S, ITS1, 5.8S, ITS2 and partial 28S)	Fw GGATCATTAACACAGTGGGCG Rv GACAGCACAGCCGGCTAG	Approx. 1000	

3.2.3 Phylogenetic analyses

Eleven *Actin* sequences and 11 ITS sequences from two populations (*Brassica* crops and *C. occulta* weeds) were obtained and deposited to the DDBJ (Table 2). All sequences were aligned and concatenated by MEGA X (Kumar et al., 2018). After concatenation, each population used was named after the sampling locations and distinguished by host source i.e., Yamaguchi-Car representing the *P. brassicae* of *C. occulta*. Phylogenetic analyses were conducted by applying the Maximum Likelihood method with 2000 bootstrap replicates to assess the reliability of phylogenetic trees. *Spongospora subterranean* f.sp. *subterranean* was selected as an outgroup, whose GenBank accession number of *Actin* and ITS is AY452193 and KF018378, respectively. DNA sequence polymorphism analysis was performed by DnaSP (Rozas et al. 2017).

Table 2. DDBJ accession numbers of Actin and ITS assigned by sampling locations and host sources (*C. occulta* weeds and Brassica crops)

	Sampling Locations	DNA markers	Accession No.
P. brassicae populations collected from *C. occulta*	Yamaguchi	*Actin*	LC663516
		ITS	LC663527
	Oki	*Actin*	LC663517
		ITS	LC663528
	Kumamoto	*Actin*	LC663518
		ITS	LC663529
	Ibaraki	*Actin*	LC663519
		ITS	LC663530
	Hiroshima	*Actin*	LC663520
		ITS	LC663531
P. brassicae populations collected from *Brassica* crops	Nagano (Chinese cabbage)	*Actin*	LC663521
		ITS	LC663532
	Kokura (Cabbage)	*Actin*	LC663522
		ITS	LC663533
	Hagi (Chinese cabbage)	*Actin*	LC663523
		ITS	LC663534
	Fukutsu (Broccoli)	*Actin*	LC663524
		ITS	LC663535
	Fukuoka (Cabbage)	*Actin*	LC663525
		ITS	LC663536
	Ibaraki (Chinese cabbage)	*Actin*	LC663526
		ITS	LC663537

3.2.4 The ITS2 secondary structure prediction

The ITS2 sequences of all populations were submitted to the ITS2 Database phylogeny workbench (http://its2.bioapps.biozentrum.uni-wuerzburg.de). Structures were predicted by either direct fold (energy minimization) or homology modelling. Prediction results of homology modelling were selected with best hit of the highest E-value and helical transfer. Other secondary structures also were retrieved from the ITS2 database using as parallels.

3.3 RESULTS

3.3.1 Sequence Variation of *P. brassicae*

Sequencing of the *actI* and *ITS* of *P. brassicae* Japanese populations (11 strains) resulted in two distinctive genotype groups, *i.e. P. brassicae* populations collected from *Brassica* crops (crops group) and *P. brassicae* populations (*Cardamine* group) from the weed *Cardamine occulta* that are homogeneous for each group with only a few intra-individual single-nucleotide polymorphisms. The sequence analysis results from the (*Cardamine* group) are remarkably different not only to the other Japanese *P. brassicae* populations but to the rest of global strains published.

3.3.1.1 *ActI* gene sequencing analysis

An approximately 700-bp *actI* fragment of the 11 *P. brassicae* strains was sequenced and analyzed. Multiple alignments performed between these *P. brassicae* strains and those retrieved from GeneBank data (including 3 *Actin* forms) reveals the diverged variation of the *P. brassicae* Japanese populations compared to global strains. While the *P. brassicae* sequences of the crops group are more similar to those of the global strains, i.e., possessing mostly silent mutations varied by *actin* forms, those strains of *Cardamine* group acquired more

noticeable mutations (Table. 3). The *actI* gene sequences of *Cardamine* group contained a total of seven point mutations exclusively representative for this group compared to the global strains, i.e., three silent and four missense mutations which are all transversion. Analysis on population diversity showed that *Cardamine* group accounted for 22% divergence compared to the entire groups, 13% among Japanese groups. This is a high ratio for a conserved coding region as *actI* gene.

Table 3. Total seven point mutations of *actI* gene sequences acquired by the *Cardamine* group compared to *Brassica* crops group, four missense mutations and three nonsense mutations illustrated in codons and translated amino acids accordingly.

		Missense mutation				**Nonsense mutation**		
		92*	104	149	153	65	95	596
P. brassicae	Cardamine group	Ser GTG	Asn AAC	Val GTC	Ala GCG	Gly GGT	Gly GGT	Ser TCA
	Brassica crops group	GCG Ala	AGC Ser	ATC Ile	ACG Thr	GGC	GGC	TCG

***** Number indicates the positions of the mutated nucleotide (underlined letter) in *Actin I* gene assigning based on the published sequence of *Actin I* (AY452179) (Archibald et al. 2004).

3.3.1.2 ITS sequencing analysis

ITS sequencing results are consistent to the *ActI* results that the *P. brassicae* Japanese population genetically separated in two groups. ITS products including small ribosomal subunit 5.8S sequence, consist of approximately 470 bp. Alignment analysis reveals a striking dissimilarity between two groups due to myriad of variations such as small-scale insertions/deletions and single polymorphisms. In total, single polymorphic sites were detected more than 11 bases for both groups. Longest insertion was found up to 11 bases observed in ITS2 region of the crop group. For the *Cardamine* group insertions were detected ranging from three to nine bases, deletions also occurred frequently from one to maximum five bases (Table 4).

The divergence polymorphic degree between the *Cardamine* group and the crop group is over 33%. Both groups were more diverged than the other groups even though the sequence of the crop group shared similarities with those of certain Korean strains. The *Cardamine* group remains the most variable one among all the groups. For the *Cardamine* group, its ITS1 region is less diverged than ITS2. In comparison to the crop group, the degree of divergence differed from 7.48% for ITS1 and higher with ITS2 at 13.5%. Only the 5.8S remains conserved in the entire groups.

Table 4. Examples of three variable sites (insertion/deletion and single polymorphism) within the ITS regions of *P. brassicae* Japanese populations (the crop and *Cardamine* group) in comparison with global strains, representing Asia (KF129413, KX430465, KX430462), North America (KY628956, KY628966), and Europe (KX011115, KX011135).

	ITS1		ITS2	
P. brassicae	32 37	375 380	441 465	
KX011115	TAT------CCA	TCGTG----------C	AGATCGACACACACACACCAAAGA	
KX011135	TAT------CCA	TCGTG----------C	AGATCGACACACACACAC-----A	
FUKUTSU	TAT------CCA	TCGTGCGCGCATTGTTC	AGATCGACACACACACAC-----A	
IBARAKI	TAT------CCA	TCGTGCGCGCATTGTTC	AGATCGACACACACACAC-----A	
HAGI	TAT------CCA	TCGTGCGCGCATTGTTC	AGATCGACACACACACAC-----A	
KOKURA	TAT------CCA	TCGTGCGCGCATTGTTC	AGATCGACACACACACAC-----A	
FUKUOKA	TAT------CCA	TCGTGCGCGCATTGTTC	AGATCGACACACACACAC-----A	
NAGANO	TAT------CCA	TCGTGCGCGCATTGTTC	AGATCGACACACACACAC-----A	
KX430465	TAT------CCA	TCGTGCGCGCATTGTTC	AGATCGACACACACACAC-----A	
KX430462	TAT------CCA	TCGTG----------C	AGATCGACACACACACAC-----A	
AB094983	TAT------CCA	TCGTG----------C	AGATCGACACACACACAC-----A	
AB094982	TAT------CCA	TCGTG----------C	AGATCGACACACACAC-------A	
HIROSHIMA-Car	CATCCCAACCCC	CCGTG----------C	A----GACACACACACACCAAAGA	
YAMAGUCHI-Car	CATCCCAACCCC	CCGTG----------C	A----GACACACACACACCAAAGA	
OKI-Car	CATCCCAACCCC	CCGTG----------C	A----GACACACACACACCAAAGA	
IBARAKI-Car	CATCCCAACCCC	CCGTG----------C	A----GACACACACACACCAAAGA	
KUMAMOTO-Car	CATCCCAACCCC	CCGTG----------C	A----GACACACACACACCAAAGA	
KY628956	TAT------CCA	TCGTG----------C	AGATCGACACACACACAC-----A	
KY628966	TAC------CCA	TCGTG----------C	AGATCGACACACACACAC-----A	
KF129413	CATCCCAACCCC	CCGTG----------C	A----GACACACACACACCAAAGA	

3.3.2 Phylogenetic analysis

Concatenated sequences data of the actI and ITS regions was applied to reconstruct the phylogenetic relationship among *P. brassicae* populations from Japan and other worldwide strains. The reconstructed ML tree (Fig. 9) clustered the crop group (six strains) with all the other cosmopolitan strains and was clearly distinct from its close relative species S. subterranean f.sp. nasturtii (outgroup). The distribution of the distinct Cardamine group is extremely skewed. It formed a well-supported monophyletic group (80% bootstrap replicates). We also ran the phylogenetic test only for the Japanese populations. The bootstrap percentage for the Cardamine clade was higher with 96% (unpublished). Only one strain KF129413 CD-1 (China) of the ITS region was also clustered in the Cardamine group clade based on the alignment pairwise of the ITS regions.

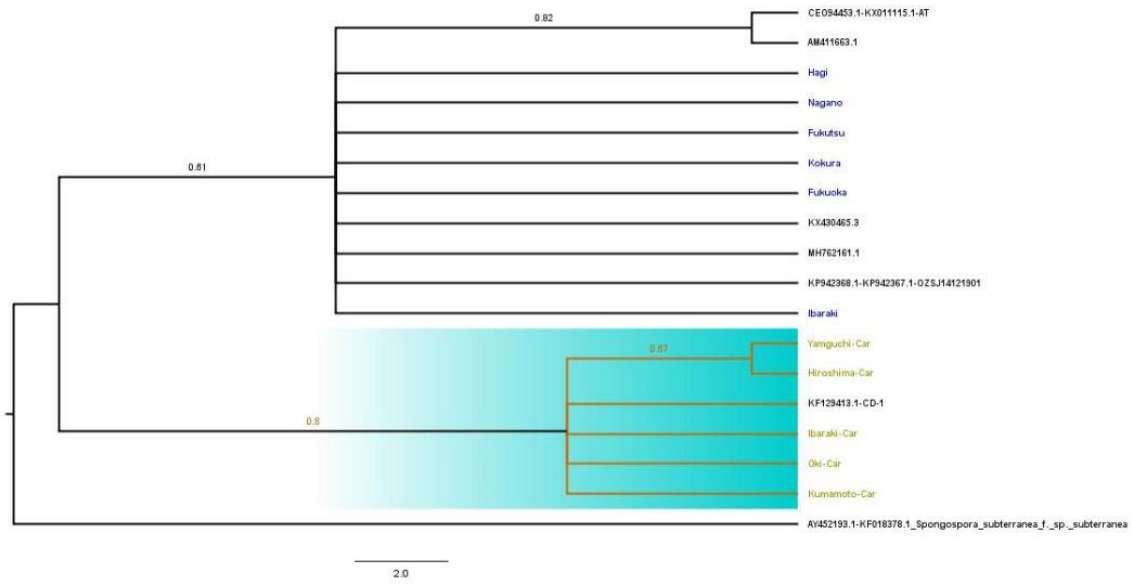


Figure 9. Inferred phylogeny of the *Plasmodiophora brassicae* populations from Japan and other global strains retrieved from GeneBank. The evolutionary relationship between the crop group and the *Cardamine* group was inferred using the Maximum Likelihood method based on

the concatenated ITS and *actI* sequences. The bootstrap consensus tree inferred from 2000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The Japanese populations named after locations are color-marked in two different colors blue representing for the crop group (Hagi, Nagano, Fukutsu, Kokura, Fukuoka and Ibaraki) and yellow for *Cardamine* group (Yamguchi-Car, Hiroshima-Car, Ibaraki-Car and Kumamoto-Car), other strains represented by retrieved GenBank accession numbers.

3.3.3 ITS2 Secondary structure analysis

The secondary structure of *P. brassicae* has been already modeled on ITS2 database based on ITS2 Workbench (http://its2.bioapps.biozentrum.uni-wuerzburg.de/). The secondary structure of cosmopolitan strains such as the AT isolate (KX011115) and many others were modeled and resulted in the unified structure (Fig. 10E) due to their similarity in sequences, as well as being similar to those ITS2 structures commonly reported on the ITS2 database website. Likewise, the overall conformation of ITS2 module for Japanese crops group fell into that group which directly folded a universal structure comprised of a central ring and four usual branched helices (helix I, II, III, and IV) (Fig. 10). However, the Cardamine group (Fig. 10A) formed a principally different structure based on its greater genetic variations acquired, segregating this group from other Japanese crops group and the rest of cosmopolitan strains. The Cardamine group in fact was the most structurally distinguishable group among those retrieved from the ITS2 database. Only two isolates from China (AF353998) (Fig. 10C) and Japan (AB094983) (Fig. 10D) shared similar loop size, the isolate from China (AF353998) (Fig. 10C) for terminal loop III and Japan (AB094983) (Fig. 10D) for terminal loop IV, respectively. The closest hit prediction for Cardamine group structure was selected based on homology model of (Fig. 10D), matched with 80.4% similarity in total, and 91.6%, 87.5%, 95.4%, and 77.7% for each transfer

helix I, II, III, IV, respectively. Other strikingly different structure is the central ring that was shaped in one circular core as opposed to double rings in the crop groups. In addition to the different base ring, helix I and III had extra loops and bulges (generated by the unpaired nucleotides) (Fig. 10A, blue arrows) as well as the sizes and positions of other loops from the Cardamine group are in complete difference with the remaining groups.

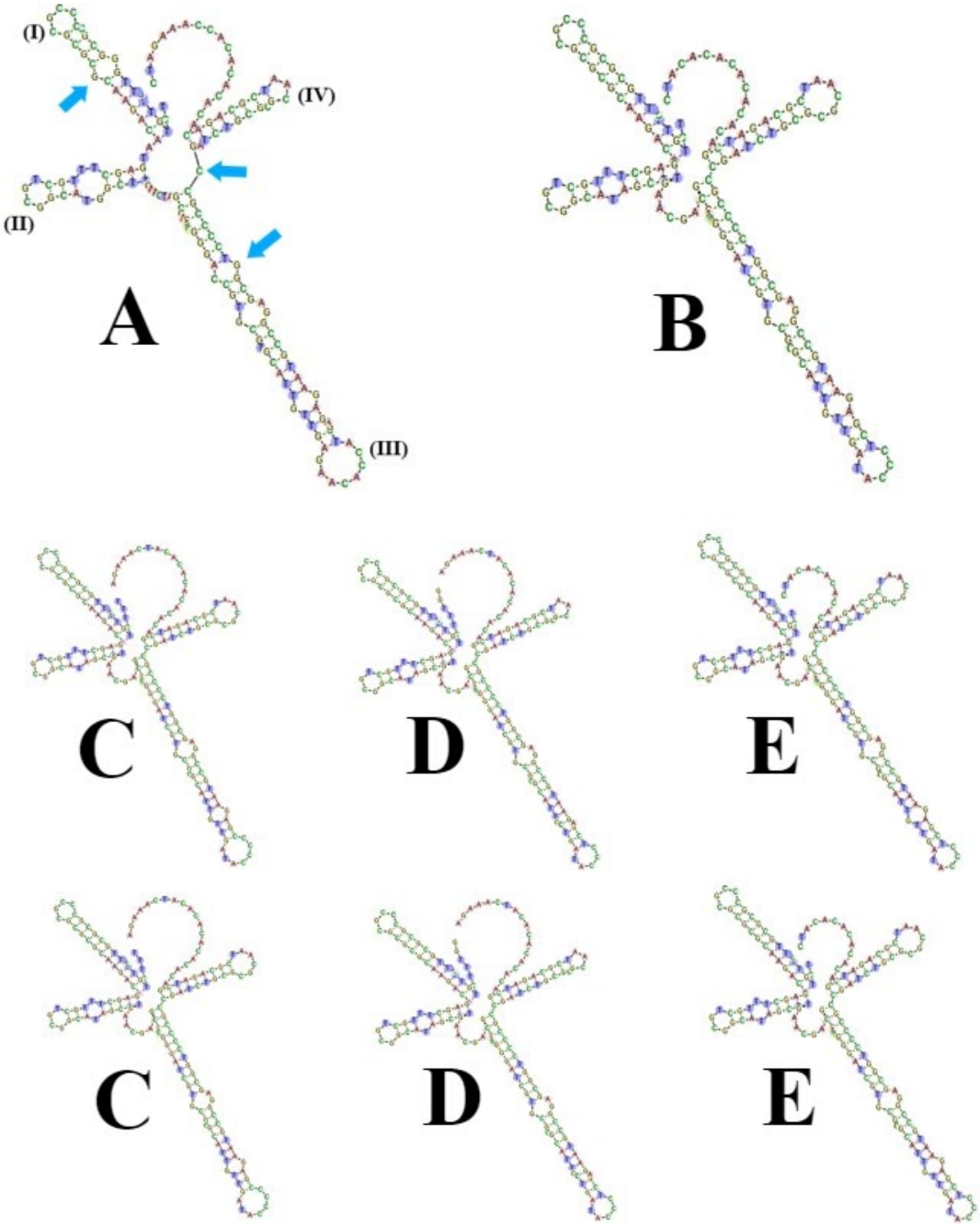


Figure 10. The ITS2 secondary structure of *Plasmodiophora brassicae* modeled from various groups. Four helices coded by Roman letters (I, II, III, and IV), colored arrows indicating visibly changes of the *Cardamine* group compared to the other groups. Alphabet letters represent individually each of the *P. brassicae* groups. **A:** the *Cardamine* group; **B:** the Japanese crop group; **C, D, E:** AF353998, AB094983, and KX011115 retrieved from ITS2 database, respectively.

3.4 DISCUSSION

Our focus in this study was the genetic diversity, as well as disparity among the *P. brassicae* populations in Japan, more importantly the aberrant *P. brassicae* population of the cruciferous weed *C. occulta*. In our case, the results reveal a strong genetic relationship between the Japanese crops and Korean groups (based on the ITS region similarity). These countries mutually share the historical and geographical backgrounds. For clubroot disease on *C. occulta*, our theory is that *P. brassicae* could be identified as a different ecotype or subspecies of the cosmopolitan *P. brassicae* due to co-evolutionary processes based on historical, observational, and more evidently the genetic data of divergence and pathological characteristics

3.4.1 Coevolutionary Relationship between *P. brassicae* and *C. occulta*

The genetic investigations on the Japanese *P. brassicae* populations fell in the same narrative that underlines the genetic makeup of the *P. brassicae* infecting *C. occulta* group deviated from the mainstream *P. brassicae*. This is also the first genetic investigation into the *Cardamine* group using the genetic markers to differentiate it from the crop group. On the one hand, the purpose of selecting *Actin* genes is because they represent one of the most conserved markers involving in morphology and other significant cellular functions. As a household gene, the *actin* genes play a vital role in cytoskeleton and several processes i.e., cell motility essential for parasitic protists as *P. brassicae* in the root invasions (Guillén et al. 2010). In this study, the *actin* genes of the *Cardamine* group harbored several missense point mutations, acting as evidences of morphological and anatomical changes of the *Cardamine* group which are difficult to be observed in an obligate microorganism as *P. brassicae*. On the other hand, the ITS regions have been widely applied for phylogenetics and molecular differentiation between closely related species due to its fast-evolving

nature (Alvarez and Wendel 2003). Although a few studies have utilized the ribosomal DNA sequences including the ITS regions to distinguish the *P. brassicae* populations geographically (Niwa et al. 2011, Laila et al. 2017), none was in success because of very few genetic variations detected as well as the similarities of the rRNA regions among the parasite *P. brassicae* and other free-living protists in the soil (Schwelm et al. 2016). Such studies have targeted on those strains infecting the *Brassica* crops, our study proved otherwise for the *Cardamine* group, possessing the largest genetic variations, especially with the ITS2 regions due to the rapidly evolutionary characters of this group as opposed to the crop group.

To provide higher resolution, ITS2 structure analysis offers more robust morphometric characters of the secondary structures known as the ring-pin model or four helix model (Coleman 2003). ITS2 is considered more conserved than ITS1 is, and ITS2 secondary structures reveal more reliability and accuracy of evolutionary characters among species since the tertiary structures play an important role in processing of precursor RNAs (pre-RNAs) (Zhang et al. 2020). The newly structured ITS2 of the *Cardamine* group had profoundly distinct characters of additional loops, bulges and different sizes of terminal tips, particularly in the helices II and III, considered as more stable and conserved than helices I and II (Coleman 2003). Even the most conserved component which is the terminal bulge of helix III of the *Cardamine* group was much different and larger than the other groups (Fig. 10). This might be a sign of a biologically different rank in contrast to the cosmopolitan group; however, additional genetical and morphological data on this group is required for further comparative analysis. The mutualistic relationship of the host *C. occulta* and pathogen *P. brassicae* also offers another interesting perspective relying on trade-off hypothesis wherein the parasite is avirulent to keep the host as a source (Alizon et al. 2009).

Plasmodiophora brassicae causes devastating disease for the cole crops. The affected crops ultimately show signs of strictly decreased growth or even premature death (Gahatraj et al. 2019). However, in the circumstance of *C. occulta*, this weed seems more tolerable to the pathogen *P. brassicae* because the biotrophic *P. brassicae* brings certain benefits to the host weed via plant hormones modification. During gall formation, myriad of hormones (salicylic acid, auxin, and many others) (Ludwig-Muller et al. 2009) are released but such metabolic modification is more likely to assist the weed in development, e.g., adventitious roots and stolons outgrowing from galls and dispersal tactic, e.g., early signs of bolting, flowering, and seedling. The latter has been observed in several weed populations growing in the nature as well as in the greenhouse experiments (data not shown). Such characteristics significantly increase the survival chances in the wild where nutrition, light sources, and land occupation are always in harsh competition within the weed community. The nature of this relationship eventually shifts from a mutualistic to a host–parasite coevolution on the account of the *Brassica* crops cultivation that is conducive to the pathogens by supplying nutrients and energy to the hosts (Dixon 2014).

3.4.2 The crucial role of the novel group *P. brassicae* to the clubroot epidemiology

To date, the publishing of genomic information (Schwelm et al. 2015) sets a milestone for many breakthrough studies that are essential for understanding several molecular aspects of host-pathogen interaction such as effectors, sexual reproduction, metabolic, transcriptomes, and secretome; however, a number of questions have arisen in its pathogenicity and physiological races remain unanswered (Schwelm 2016). One of the factors significantly contributing to pathogenicity is the interactions between wild plants and pathogens is often understudied. In fact, the disease dynamics is directly linked to population genetic structure of the pathogens. This is critically determined by host diversity

and disease pressure (McDonald and Linde 2002). Several molecular investigations on diversity among the pathotypes and the relevant physiological races mainly focus on differences of these pathotypes and races; however, the attempt to infer the pathogenesis and relevant molecular mechanisms behind the newly rising pathotypes or virulent races are subject of a long-debated matter.

Our primary concern for this newly rising "isolate/population/strain" is its vital role attributing in the pathogenicity of the cosmopolitan group because the pathogen *P. brassicae* is well-known for its amenability to accept foreign DNA carrying high mutations so that it is capable of widening the arm races, as well as host ranges (Schwelm and Ludwig-Müller 2021). Owing to the fact that the host weeds *C. occulta* possess a highly genetic diversity than any cultivated counterparts of the *Brassica* crops and this might potentially diversify the genotype of the pathogen *P. brassicae* in a sense that the sensitive weed hosts are eliminated, and highly virulent pathogen strains survived. This natural elimination process confers the novel virulent genes and beneficial mutations to the pathogen populations. Our team has been receiving reports from local farmers describing multiple clubroot outbreaks of the *Brassica* crops cultivated from converted paddy field, as tested by William's system the pathotype of the outbreak isolates resulted in less dominant pathotype 7 (where pathotype 1 and 4 found common in Japan (Kim et al., 2016)) that is able to escape the clubroot-resistant cultivars (unpublished). Although the newly introduced *Brassica* crops are clubroot-resistant, nevertheless their genetic homogeneity as a crop renders them vulnerable to *P. brassicae* populations of the wild weeds that would otherwise exist in low numbers. Another example for this phenomenon is the clubroot disease occurring on Wasabi farms (Tanaka et al. 1994), wherein these Wasabi plants (*Eutrema wasabi* Maxim.) were cultivated on drained paddy fields previously infested with highly clubroot-infected *C. occulta.*

More concerningly, the dispersal ability and invasiveness of the host weed *C. occulta* are considerably threatening. This fast-growing weed has been reported present intercontinentally, from Asia China, Taiwan, Thailand, Vietnam to Australia, and Northern and Southern America, Canada, USA, and Mexico and recently Europe where it was found affiliated in natural wetland areas as lakeshore or even dry lands along roadsides but also anthropogenic habitats, e.g., flower beds and plant nurseries (Marhold, 2016). The richness of the *Cardamine* species should greatly be taken into consideration because they are one of the largest genera of the Brassicaceae family, consisting of over 70 taxa and at least 340 species (Mandáková et al. 2019) spreading worldwide from natural ranges to man-made habitats (Lihová et al, 2006). Three species of the genus *Cardamine* have already been recorded as hosts of *P. brassicae*, namely *C. pratensis* L. (Nowicki, 1973), *C. belldifolia* L. (Nowicki, 1973), and *C. hetrophylla* O. E. Schulz (Gibbs, 1932). *Cardamine hirsuta* is also an invasive and overabundant weed in Japan, often found along roadsides and dried areas (Kudoh, 2017). As tested in the laboratory, this invasive weed seemed to be susceptible to the *P. brassicae* pathogens extracting from the *C. occulta* weed because of the higher rate of infection and visible tumor compared to the statistical data collected from the *C. occulta* weed (Tanaka and Ito 2013). Field survey of clubroot disease on *Cardamine scutata*, naturalized in Japan, usually thrives on riverbanks and mountainous wetlands, also harbored clubroot pathogens but the genetic populations were unidentified throroughly (Tanaka and Ito 2013). The likelihood of the pathogen *P. brassicae* detected with the *Cardamine* species is tremendously high. The aforementioned KF129413 CD-1 strain from China shares a nearly identical sequence to the ITS regions of the *Cardamine* group. Therefore, the presence of this strain across the *Brassica*-crop cultivation zones in China is highly predictable as well as in Korea due to the rice cultural history and geographical background mutually shared between these countries. We suggest further comparative genomic analysis should be conducted on this newly

discovered *P. brassicae* group to elucidate its major impact on shaping pathogenicity and the dynamics of clubroot disease corresponding to the coevolution of the weed host-pathogen interactions.

CHAPTER 4

THE *P. BRASSICAE* POPULATIONS INFFECTED *C. OCCULTA* REVEAL THE COMPLEX DYNAMICS OF RIBOSOMAL INTRONS

4.1 INTRODUCTION

Plasmodiophora brassicae, a plasmodiophorid (phytomyxean, rhizarian) biotrophic pathogen is a causal pathogen of clubroot disease on cruciferous species. Several studies on this obligate intracellular protist mainly focus on the cruciferous crops because of their economic importance for agriculture whereas other wild *Brassica* species often become overlooked. Our group has previously reported the clubroot disease has been identified frequently on the cruciferous weed *Cardamine occulta* (Tanaka et al. 1993). This overwinter weed is abundantly found in several human-made habitats, particularly in the well-irrigated paddy fields where this wild weed flourishes in the post-harvest periods and has since established as part of the weed community in the rice cultivation of Japan. The clubroot incidents occurring on this weed have been described from pathological traits diagnosed on fields to multiple stages of the secondary infection observed under microscopy (Lam et al. 2022a). In addition to phenotypical identification, the genetic differences acquired in *Actin* gene and internal transcribed spacers (ITS1 and ITS2) were also elaborated, in particular focus on the secondary structure of ITS2. The ITS2 structured from the *P. brassicae* group infecting *C. occulta* (known as the *Cardamine* group) were indicative of distinct genotypes as opposed to the mainstream *P. brassicae* attacking on the *Brassica* crops (Lam et al. 2022a). Likewise, we extend the genetic investigations on the other non-coding regions such as the intronic sequences established in the ribosomal DNA (Small Subunit and Large Subunit) by following a similar logic used for the ITS regions since these intronic

sequences are involved in operating the processing the rRNA precursors (Flipphi et al. 2013, Flipphi et al. 2017).

The long debate over the significant role of rDNA regions in geographical differentiation was initiated by Niwa et al. (2011). The geographical *P. brassicae* populations in Japan were reported to carry a highly polymorphic LSU that contains the novel lengthy introns. However, this study was refuted by Schwelm et al. (2016) who confirmed the pathogen *P. brassicae* has no polymorphism in the LSU regions and corrected those novel introns found in Japanese populations were not originated from *P. brassicae* but were misidentified by a chimeric PCR product from the glissomonad *Neoheteromita globose* instead. This story was succeeded by Laila et al. (2017) who reaffirmed the sequence variations detected in rDNA of geographical isolates in Korea and the SSU intron I was deleterious in several isolates. Nevertheless, Schwelm and Neuhauser (2017) were persistent on the concept of no rRNA polymorphism detected in nuclear ribosomal DNA for the plant pathogen *P. brassicae* on global scale.

In our study, we still value the role of nuclear ribosomal DNA of *P. brassicae* crediting for taxonomic classification but slightly different in the target, which we focus on the *Cardamine* group, more specifically the intronic sequences. The intronic sequences of this groups are immensely diverse, this could be attributed to the life style and genetic make-up of the neotype host weed. As indicated above, the weed hosts *C. occulta* share a completely different life style in contrast to the crop hosts. We would consider the *P. brassicae*-infected *C. occulta* as “natural populations” compared to the *P. brassicae*-infected brassica crops, i.e., natural vs domesticated. As an obligate biotrophic pathogen, *P. brassicae* are genetically compelled to shift to a new host. The genetic engineering of the *Cardamine* group might be the vital source for the genetic variations in the global *P. brassicae* populations. In this perspective, this might explain the clubroot pathogenicity and pathotypic-associated genes because these characters may stem from such variations. Therefore, this research might open

another perspective on researching population-level diversity in the *P. brassicae* via the evolution of introns, more likely bridging variations to pathogenicity and etiology of the clubroot disease.

4.2 MATERIALS AND METHODS

4.2.1 Clubroot samples collection

The clubroot samples were collected from diseased plants between the period 1994-2014 from multiple locations across Japan (Table. 2). The collected clubroot galls were washed thoroughly and stored at -40°C.

.

4.2.2 DNA Extraction

The clubroot samples were washed thoroughly by tap water and then distilled water. Tissue samples were grounded by pestle and mortar in liquid Nitrogen to extract the resting spores, described previously by Osaki et al. (2008). Centrifugation repeated at 50rpm for spore suspension until the supernatant became completely transparent. DNA was purified from resting spores based on modified methods described by Ito et al. (1997)

4.2.3 Primer designation and Sequencing analysis

All primers were designed by Primer3 software (Available online: http://primer3.ut.ee). Primers of 18S and 28S exonic and intronic regions were designed based on KX011115 in the GenBank database (Table 5). Due to intron loss and gain pattern (the first intron in 18S region of the crop group and several introns in the *Cadamine* group), extra primers were designed to target the exonic regions or related landmark regions (ITS regions) in order to ensure the desired intronic regions belonging to *P. brassicae* (as in many others free-living or parasitic protists such as *Woronina pythii* Goldie-Smith or *Spongospora subterranea* f. sp. *Subterranean* have similar Ribosomal sequences). These extra

primers were designed in similar PCR conditions, thus were able to match randomly with other primers to vary the PCR products that including exonic parts or landmarks like ITS regions, which are exclusively identifiable for the pathogen *P. brassicae*. All sequences were deposited to Genbank with information on varieties and sampling year (Supplementary Table 1). All sequences were aligned and analyzed by MEGA X (Kumar et al. 2018)

Table 5. List of primers designed for ribosomal exonic and intronic regions with PCR conditions, products size and targeted regions Primers

Sequences		**Forward (Fw) and Reverse Primers (Rv)**	**Products**	**PCR Conditions**
18S (Small Subunit)	Intron 1	Fw TATCGAGGATCCATTGGAGG Rv TGTCGTAACGCTCTTTTGAA	530	96 C 3 min, 96 C 30s, 56-58 C 45 s 72 C 1 min, 32 cycles, 72° C 10 min
	Intron 2	Fw TCGCAAGGCTGAAACTTAAA Rv CTCTACATCTCTGGACCTGG	502	
	Intron 3	Fw TGGGCTCTTAGAAGAAGGAG Rv GCAACTTGCGTTCAAAGATT	800	
	Exon	Fw TTGATCCTGCCAGTAGTGAT Fw ACCTATTATCGAGGATCCATT Rv CACACAAGTCCAACTACGAG Rv TGTGTATGTATGGAACTTCAG		
28S (Large Subunit)	Intron 1	Fw TGAAGAAATTCAACCAAGCG Rv ATGCGCGTCACTAATTAGAT	595	
	Intron 2	Fw TCAATGTGGAAATTTCACACT Rv ATCGAAGAATCAAAAAGCAAC	592	
	Exon	Fw GATTACCCGCTGAATTTAAGCA Fw AACCTATTATCGAGGATCCATT Fw TAATCGCCCTGGGTATGGTA Rv AGTTACTCCCGCCGTTTACC		

4.2.4 Sequence Annotation

The MFannot tool (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl) was used to identified introns as group I or group II introns as well as open reading frames (ORFs) (Supplementary Figure 8). Annotation of ORFs was reviewed and revised by Rfam (https://rfam.org/) and BLAST homology searches against the NCBI protein database.

4.2.5 Secondary structure prediction

The secondary structures of the intronic regions from the *P. brassicae* of the *Cardamine* group were formed by The UNAFold Web Server (http://www.unafold.org) to test for Thermodynamic Structural Entropy (ΔG) (Supplementary Figures 1-7). The entropies and folded secondary structures of the *Cardamine* group were highly possible in comparison with the entropies and structures folded in the crop group.

4.3 RESULTS

4.3.1 The Pattern of loss/gain introns distributed within the *P. brassicae* populations in Japan

Based on intronic distribution established in the ribosomal DNA of the AT isolate (KX011115) (Fig. 11), the pattern of rRNA intronic distribution among the *P. brassicae* Japanese populations was described in Table 6. Of all five introns, the gain and loss pattern only occurred specifically in the 18S intron I of the crop group whereas the *Cardamine* group appeared to suffer massive intron losses in both SSU and LSU. The *Cardamine* group only possessed intron I either in the 18S or 28S, except for the Nara population harboring both introns I in 18S and 28s regions. The other introns of the *Cardamine* group (18S intron II, III and 28S intron II) were in complete absence (Table 6).

Table 6. The Absence-Presence Pattern of Intronic Sequences distributed in Ribosomal DNA genes (18S and 28S) among the *P. brassicae* populations infecting *C. occulta* (the *Cardamine* group) vs. the *Brassica* crops.

Intronic sequences		The *Cardamine* group	The crop group
18S		Variable	Variable
	I		
	II	No	Yes
	III	No	Yes
28S	I	Variable	Yes
	II	No	Yes

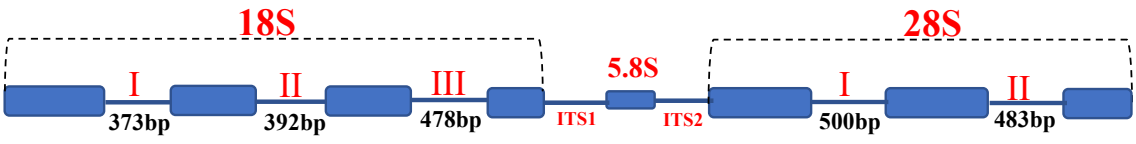


Figure 11. Sizes and distribution of the introns in the ribosomal DNA of *P. brassicae*. Shaded boxes represent the exonic parts connected to lines with Roman numerals (I, II and III) above that present for intronic sequences for each region 18S and 28S (dash lines) of the rRNA.

The BLAST results from the *Cadarmine* group have shown certain degree of similarity with *Woronina pythii* or *Spongospora subterranea* f. sp. *subterranean* sequences due to the status of ribosomal intronless sequences, however when aligned and evaluated the exonic parts of the *Cadarmine* group shared more similar identity to the sequence of the *P. brassicae* AT strain.

4.3.2 Intron Analysis of Small (18S) and Large (28S) Subunit rRNA genes

4.3.2.1 Small subunit (18S) rRNA gene

A completely deleted 18S intron I was found in many populations. However, both groups also had heterozygous 18S intron I sequences (lane 6, 11, 13, 14, 15, 16, 17, 18, and 19) that harbor both intron-free and intron-bearing within one single field population (Fig. 12). This result is consistent with the research conducted by Laila et al. (2016). The authors reported that *P. brassicae* crop group in particular for several Korean populations also shared the common pattern of retention and loss solely occurred in the 18S intron I sequences.

As Schwelm et al. (2016) addressed, studies related to rRNA genes of *P. brassicae* might be misidentified with other glissomonads *Neoheteromita globosa* or *Spongospora*-like plasmodiophorids which are also living overabundantly in the soil. Therefore, primers were exclusively designed to target the exons from those intron-poor populations and sequencing results affirmed the exons the investigated populations belonging to the *P. brassicae*.
Sequencing results of the 18S introns I indicated most of the crop group populations remained conserved in sequences and sizes. They shared the same length 374 bp with the cosmopolitan isolate AT's. Three populations possessed a haplotype illustrating by the transition point mutation A/G in 5' upstream position near the flanking exon-intron boundaries, but the sequenced introns from the rest were identical to the AT sequence (Fig. 12, position 24). The two

populations Hagi and Itoshima were separated from the crop group by the sizes and mutation acquisitions, particularly with the Itoshima population. The size of 18S intron I in the Itoshima population was significantly longer (407 bp) (Fig. 12, position 374) due to three insertions at the specific sites. These specific sites may be related to the intronic secondary structures (branching site, binding site and lariat structures) in the self-spliceosomal activity. The Hagi population had fewer mutations than the Itoshima's but all of other intronic sequences from the SSU to the LSU, were detected with high level of polymorphic contains.

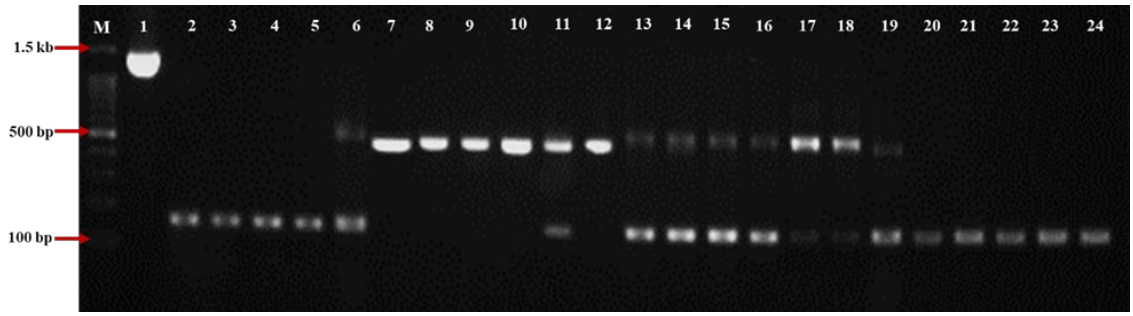


Figure 12. PCR products illustrate the 18S intron I distribution (374 bp) of two groups, the *Cardamine* and crop group. Lane M represents the standard ladder (arrows pointed at 1.5kb, 500bp and 100bp). Lane 1: *Rorippa palustris* (906 bp). Lane 2-7 represent the *Cardamine* group including Kumamoto, Yamaguchi, Akiota, Tochigo, Nagasaki and Nara, respectively. Lane 8-24 represents the crop group including Fukuoka, Aichi, Itoshima, Hokkaido, Ishikawa, Gunma, Nagasaki, Onga, Ibaraki, Saga, Hagi, Osaka, Fukutsu, Shimane, Kurume, Wakamatsu, and Karatsu, respectively. Populations distinguished by genotypes such as homozygous either missing the introns displaying the 102 bp of the exonic parts (2-5, 20-24) or retaining one band of the introns (7-10, 12), heterozygous by having both bands (11, 13-19). Faint bands observed in several lanes, other than the distinct and sharp main products due to the purity of DNA.

The *Cardamine* group led a different pattern in sequences and acquired mutations as opposed to the crop group. One special inclusion to the *Cardamine* group is the *P. brassicae* populations collected from the wild weed *Rorippa palustris* (LC716109) due to its 18S intron I homologous to those of the *Cardamine* group, but its intron was longer in sequence by carrying an encoded-ORF (approximately 550 bp) (Supplementary Figure 7). While four populations (Yamaguchi, Kumamoto, Akiota and Tochigo) were in devoid of all introns in 18S region, two populations Nara and Nagasaki carried the 18S intron I and their sequences displayed homology in term of sizes and polymorphisms acquired. The *Cardamine* group had several deletions outnumbering the insertions, resulting a shorter size (359 bp) compared to a standard size 374 bp of the 18S intron I. These polymorphic InDels appeared at the similar positions wherein occurred the mutations of the Itoshima and Hagi populations (Fig. 12). The *Cardamine* group also had the intronic upstream deletions and nucleotide shifting as well as transversion point mutations at the downstream where flanking regions including the splicing sites start. The 7-nucleotide cluster of the *Cadarmine* group positioning at 150 site (Fig. 13) was mutated (T**A**T**A**TG**T**/TTTTTGA-*Cardamine*/crop AT, mutations described in bold letters), mimicking the flanking in the downstream (positioning from nucleotide 384 to 391) of the cosmopolitan isolates. Meanwhile, at that same flanking region of the *Cardamine* group wherein single polymorphism (A/T) occurred (TTT**A**TGA/TATATGT-*Cardamine*/crop). These phenomena recurred regularly when observed the polymorphisms in the introns of the 18S regions. For instance, the terminal flanking of the 18S intron 3 from the Hagi population had a single point mutation (Transversion T/C, TA**C**ACG/TATACG), this 5-nucleotide cluster was found in the midstream positions (362-366, based on AT) wherein it was mutated (**T**----G/CACACG, deleted nucleotides described in dash lines) in the Hagi population. This extended mutated cluster (AT----GTGTGT) coincided with terminal flanking of the 18S intron I from the Itoshima population (Fig. 13).

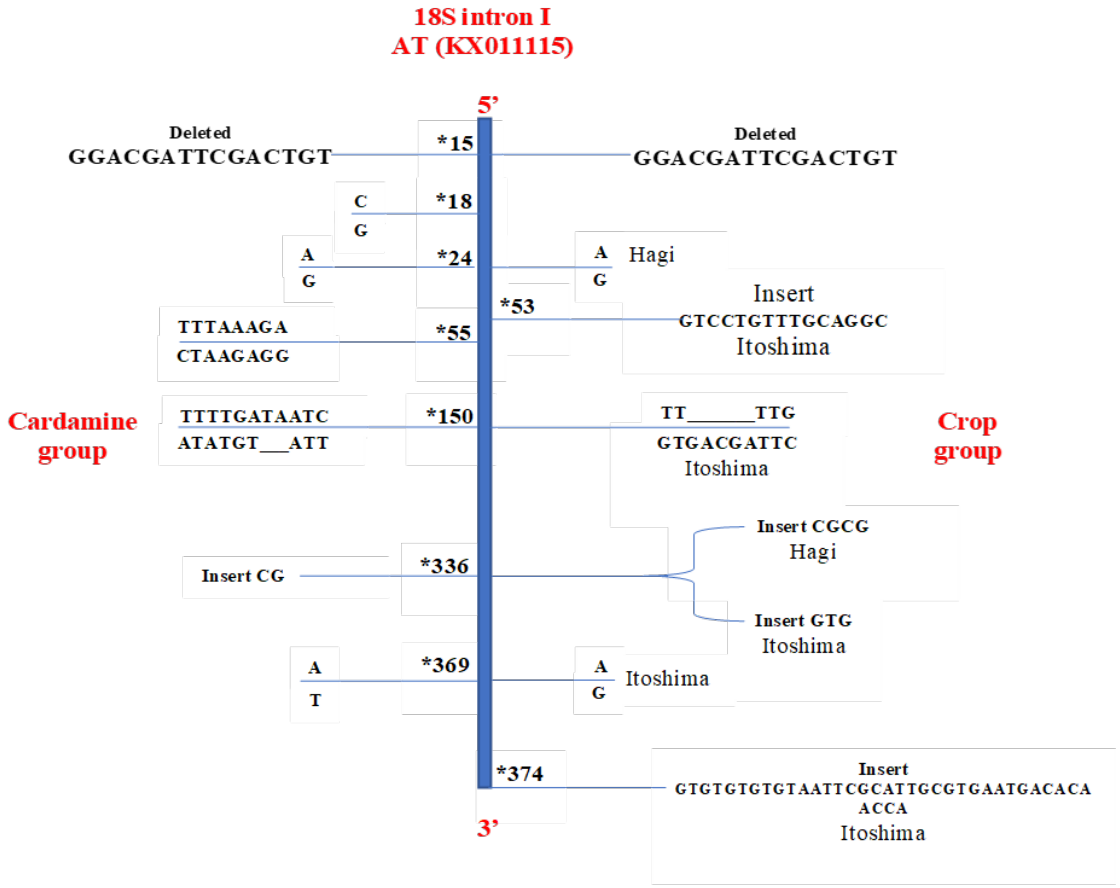


Figure 13. The scheme of intraspecific variability acquired in the 18S intron I from the examined populations of two groups the *Cardamine* and the crop group designated on the standard of 18S intron I of AT isolate (KX011115). Number with * describe the positions' nucleotide originating in KX011115. Some mutated nucleotide clusters and single-point mutations were demonstrated under the lines (where above the line for original sequence of KX011115), underline among the clusters describes the deleterious nucleotides.

4.3.2.2 Large subunit (18S) rRNA gene

The crop group remained being intron-positive with two introns fixated in the LSU region. In contrast to the crop group, the *Cardamine* group had only a single

intron 28s intron I, following the similar pattern of SSU region (Fig. 14). Of all the *Cardamine group*, only Nara population owned two introns (18S intron I and 28 intron I) from both regions SSU and LSU, other populations (Kumamoto, Yamaguchi, Nagasaki, Akiota and Togochi) possessed only one single intron for the entire rRNA genes. The 28s intron I pattern of the *Cardamine* group was more diverged in length and intraspecific variability than in any introns from 28S regions of the crop group (Fig. 15).

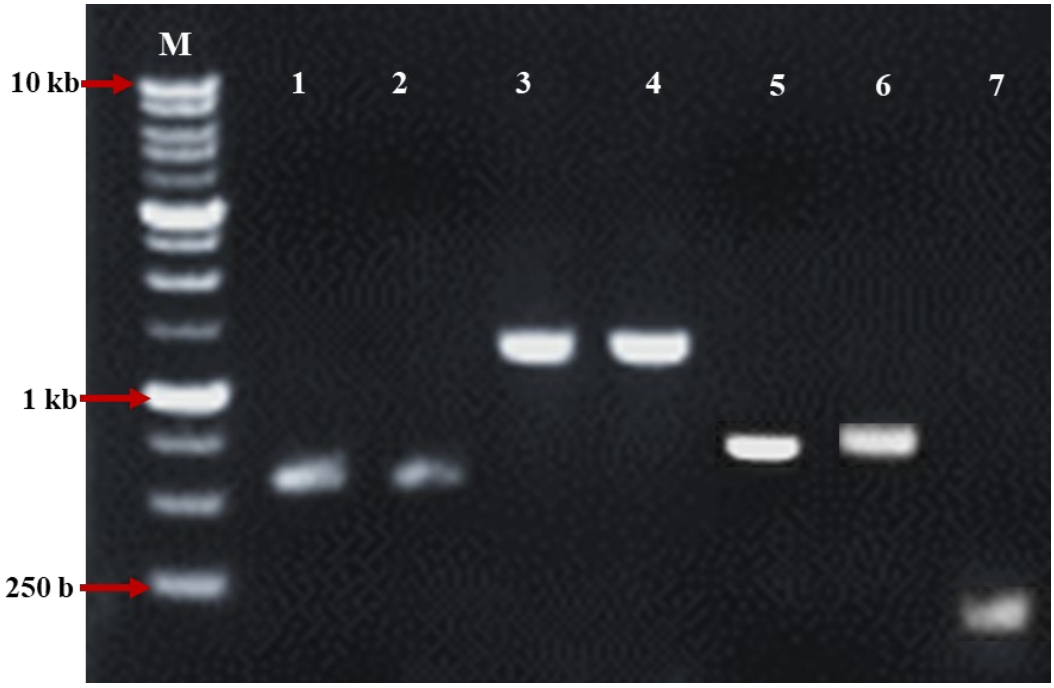


Figure 14. PCR products demonstrate the variable DNA sizes of the *Cardamine* populations. Lane M represents the standard ladders (arrows pointed at the 10 kb, 1 kb and 250 bp). Lane 1: positive control, 501 bp 28S intron I. Lane 2-7: Nara, Akiota, Tochigo, Kumamoto, Yamaguchi and Nagasaki, respectively.

Comparative nucleotide sequence analysis showed the *Cardamine* group shared a partially homologous sequence of approximately 438 bp respective to the crop group, 87.7% sequence identity between two groups. After the homologous part shared with the crop group, the *Cardamine* group had extra sequences added which make them longer introns than the 28S intron I of the

crop group) (Fig. 14). Among the *Cardamine* group, the longest intron was located in the Hiroshima populations (Akiota and Togochi) owing to the 766 bp additional sequence harboring a unique ORFs (identified by MFannot), while the shortest intron (561bp) was located in the Nara population. The extra sequences harbored in all of the populations (Yamaguchi, Nara, and Kumamoto) were greatly varied by size but commonly showed tracts or repeative elements similar to the extra sequences of the Hiroshima populations. For instance, the Kumamoto population had the same downstream sequence (77 bp) of the Hiroshima populations, bearing a repeative element (CA tracts).

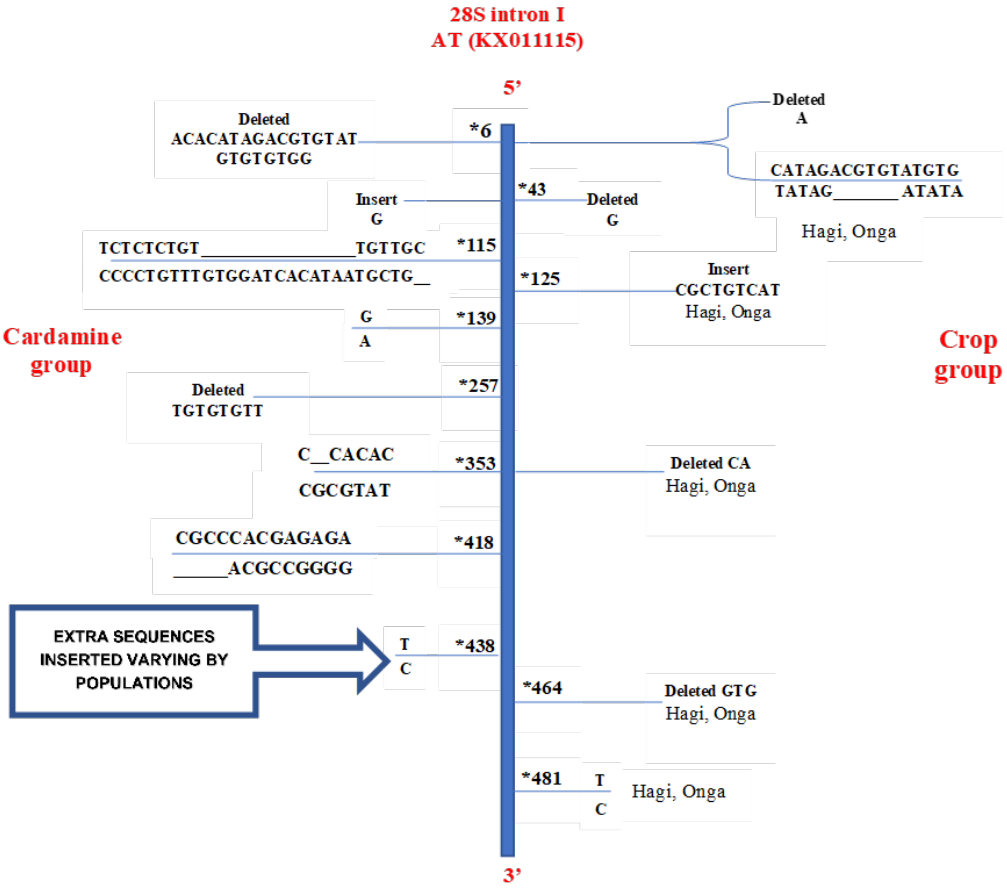


Figure 15. The scheme of intraspecific variability acquired in the 28S intron I from the examined populations of two groups the *Cardamine* and the crop group designated on

the standard of 28S intron I of AT isolate (KX011115). Number with * describe the positions' nucleotide originating in KX011115. Some mutated clusters of nucleotides and single point mutations were demonstrated under the lines (where above the line for original sequences of KX011115), underline among the clusters describes the deleterious nucleotides.

Table 7. The polymorphisms acquired in the 28S intron II identified in the aberrant populations from the crop group (Hagi and Onga) and from the cruciferous weeds *Capsella bursa-pastoris*. Numbers describe the positions of nucleotides based on the *P. brassicae* AT isolate (KX011115) wherein the mutations occurred

	Mutated Positions identified in 28S intron II					
AT isolate (KX011115)	30 C	45 G	200 G	282 G	342-345 ATC	345-347 ATA
Capsella bursa-pastoris	G	G	A	deleted	deleted	ATA
Hagi and Onga	G	T	A	deleted	ATC	deleted

The 28S intron I of the crop group was more homogenous than that of the *Cardamine* group in the scope of size polymorphism and InDels acquisitions (Fig. 15). Most of the tested populations had the identical sequences to the AT isolate, except for a single nucleotide deleted at the 5' upstream of the exon-introns, suspected as similarly as in 18S that it might be the haplotype of the *P. brassicae*

affecting on crops in Japan. The Hagi population, as referred above, accumulated plethora of polymorphisms not only in 28S intron I but also the intron II. Another deviated population, the Onga population, also shared similar mutations acquired in both intronic sequences (28S I and II) with Hagi but had fewer polymorphisms. More interestingly, the Onga population had a point mutation (inserted a nucleotide G) at the exon flank (TACCACAGGG**G**) where intron/exon boundary meets. This "exon flank" cluster is also repeated at the 5' upstream of the 28S intron II (Table 7). It is noteworthy that one of the collected samples from the shepherd's purse weed *Capsella bursa-pastoris*, had the 28S intron II sequence homologous to the Hagi and Onga population due to similar variations attained (Table 7).

4.4 DISCUSSION

In this research, the secondary structure of the ITS2 analyzed from the *Cardamine* group stood separately from the rest of cosmopolitan groups based on the collective ITS2 database of several *P. brassicae* populations, created from ITS2 Database phylogeny workbench (http://its2.bioapps.biozentrum.uni-wuerzburg.de) (Lam et al. 2022a). The sequences of the *Cardamine* group in this research are completely novel and might be controversial on the identity and functions (see Supplementary Figures 1-7, for the foled secondary structures and entropies)

In this study, our focus was on the *Cardamine* group possessed the striking patterns of intron gains and losses in the *Cardamine* group as well as the distinctive polymorphic introns of the rRNAs among the geographical populations. The results also confirmed the original hypothesis we proposed

previously that the *Cardamine* group is a discrete group compared to the cosmopolitan *P. brassicae*.

4.4.1 Early or late theory for the intronic distribution in *P. brassicae*?

The retention and loss pattern has occurred frequently in genetic basis of various species across the eukaryotic evolution (Rogozin et al. 2012). This heated debate has waged war between the two contrasting theories of intron distribution "early vs. late" (de Souza et al. 1998). The first theory, called the early-intron that hypothesizes the ancestral genomes had contained the DNA segments which played a vital role in protein synthesis, and during the evolutionary process these segments were completely removed in prokaryote lineages but retained scatteringly as intragenic sequences known as introns in eukaryotic biology (Koonin 2006). The counter point-of-view to the early-intron theory argues that introns are vestigial, originating from parasitic DNAs that have inserted to the host genome through the evolutionary host-parasite relationship, the introns thus are obliged to splice themselves out during the protein-synthesizing in an effort to retain in the host genome, a survival tactic compromised with the host genome (Belshaw and Bensasson 2006). Over the years of this long debate, these hypotheses have complemented from one to another, resulting in a weaker version of each theory (Belshaw and Bensasson, 2006). The ribosomal RNAs of the *P. brassicae* populations varied by two groups (the Brassica crop and Cardamine group) can be taken as an example for this circumstance. For the first theory, several populations of the crop group (from this study and Laila et al. 2017) reported having lost 18S intron I that is the first sign, and more increasing body of evidences come from the Cardamine group whose introns in the rRNA have been gradually eliminated in large quantity. On the counterpoint, the ribosomal RNA of the pathogen *P. brassicae* has traces or remnants that facilitate the intronic insertions that corroborate the late intron theory. As observed in the ribosomal introns of the AT isolate (KX011115), few introns had conserved

tracts (7-10 nucleotides) that flank the upstream of subsequent exons. For instance, the 18S intron III had 9 nucleotides (TTCCGTAGG) that flank the last exon of SSU; likewise, in the LSU region, the 28S intron I and II had 7-nucleotide (CTCTTAA) and 10-nucleotide (TACCACAGGG) sites, respectively that repeats the successive exons. These tracts, known as "proto-splice sites," consider as relics from the signaling sites to initiate the intronic insertion (Long and Rosenberg, 2000).

In addition to the recognition sites, the intronic ORFs harbored in the 28S intron I of the Cardamine group, might perform as transposable-like elements or also known as retrotransposons that insert themselves to the intronless sequences and genes. This phenomenon might possibly explain the abundance of intron quantity in the Brassica crop group due to the copy-paste mechanism operated by these putative retrotransposons. To understand the insights of the massive intron loss and gain occurring in the Japanese *P. brassicae* populations and answer the question of which intron models suit for the case of *P. brassicae* introns, comparative genomic analysis should be deployed on population-genetic diversity of both groups. The main target is the Cardamine group which has been proven having a highly diversified genetic materials as well as the intron distribution and variability. Indeed, only five populations were tested in this research, yet existed three patterns of intronic distribution for the Cardamine group (either only one intron harbored in SSU (Nagasaki) or LSU (Kumamoto, Akiota and Togochi from Hiroshima, Yamaguchi) and having two for each region (Nara). It is also worth mentioning an exceptional case that one *P. brassicae* population extracted from the crop group was tested intron-negative for the entire rRNA regions. This case might appear as a contamination of DNA extraction as reported by Schwelm et al. (2016). Indeed, the nature of the parasite–plant relationship renders the DNA purification difficult, even more difficult for the wild weeds due to the contamination of other free-living protists, unidentified pathogens and more laboriously, the tasks of culturing the weeds.

Nevertheless, the possibility of being intron-void rRNAs for the Cardamine group is highly predictable based on the intron early theory that advocates the phenomenon of intron loss for the entire genome in the eukaryotic evolution (Roy and Gilbert 2005, Roy and Hartl 2006, Roy and Penny 2007). This phenomenon is actually not uncommon, in particular for a unicellular obligate parasite as *P. brassicae*, if taken its closely relative *Spongospora subterranea* f. sp. *Subterranean*, also known as biotrophic pathogens, as empirical evidence for being intron-poorer than *P. brassicae* as well as intron-negative in the rRNA regions (Stjelja et al. 2019).

4.4.2 The intronic evolution equates to speciation event of *P. brassicae* populations?

We have provided the evidences of the phenotypic traits (characteristic symptoms and microscopic evaluations) combined with the genetic makeup (ITS and *Actin*) of the Cardamine group and raised the question about its taxonomical rank (Lam et al. 2022a,b). That question still remains unanswerable because the fine line to define the limits of a new species is controversial. The goal of this study to illustrate the speciation events from the perspective of intron evolution, especially for Cardamine group via the scope of the early vs. late theories. The stark contrast between Cardamine group and the crop group has proven the evolutionary dynamics of intron occurred in the *P. brassicae* populations. The mechanism behind the intron dynamics might imply the correlation or association with the event of the host shifting (or host-switching) from the domesticated crops to the wild weeds.

Although *C. occulta* is a crucifer, its genetic materials, life cycle and conditions related to growth and reproduction are entirely disparate to those crucifers of the crops. During the process of domestication whereby the genomes of the *Brassica* crops have been artificially modified by selecting the desired

genes (Flint-Garcia 2013). By contrast, the genome of *C. occulta* weeds, considered as the most dispersible of *Cardamine* species (Al-Shehbaz et al. 2006), accumulates plethora of genes and genetic variations that are able to confer the adaptative advantages to variety of niches (Šlenker et al. 2018, Rutland et al. 2021). These driven forces generate an evolutionary divergence from several scopes of a host-parasite relationship. The host-parasite dynamics requires alteration of both genotypes and phenotypes alternatively between host and pathogen (Ebert and Fields 2020), meaning that the pathogenic genome might be synergistic with its hosts to increase the survival opportunity (Märkle et al., 2021). In this scope, the genetic variations acquired by the pathogens infecting the domesticates might not be as enriched as those acquired infecting the wild weeds due to the less diversified genome of the domesticates.

Schwelm et al. (2016) stated the rRNA sequences of *P. brassicae* were not proper candidates for isolates or pathotypes differentiation due to no genetic variation detected in the LSU region. We concur partially with this statement for most of the cosmopolitan isolates attacking the Brassica crops. On the contrary, we also counter that point of view by demonstrating the *P. brassicae* colonizing the *C. occulta* weeds whose genetic diversity is much of otherwise. The octoploid genome of the weed (2n=8x=64) is the product of double duplication event (Mandáková et al. 2019) which can be construed as an advantageous factor for the evolutionary process and diversification (Crow and Wagner 2006, Ruthland et al. 2021). This genomic plasticity of the polyploidy host has the potential to confer new DNA elements to the parasitic recipient genome and vice versa. This suggests the underlying mechanism of intron diversification separates two groups, with additional, perhaps this disparate diversification propel the mode of speciation operated in those *P. brassicae*-infecting-the-weeds lineages.

4.4.3 Prospective research for intragenic sequences of *P. brassicae*

The genome of *P. brassicae* was analyzed as compacted by principally decreasing intergenic spaces as well as gene losses for the purpose of being entirely reliant on intracellular resources from hosts (Rolfe et al. 2016). Despite that fact, *P. brassicae* was still evaluated as an intron-rich parasite (Bulman et al. 2007, Stjelja et al. 2019), the rRNA genes that are particularly intron-dense. As hypothesized above with the host shifting event, the intron dynamics regard to numbers and patterns are segregated two groups but do the tertiary structures and functions of those introns exclusively play for the *P. brassicae* lifestyle, for instance involving with the parasitic or survival strategies?

The ribosomal introns are considered more diverse and dominant due to their functions essential for gene regulations, in particular for ribosomal protein genes in gene expression and diversification of the protein repertoire (Rogers, 2019). The intron prevalence of the crop group may reflect the up-regulated genes and protein biosynthesis due to stability of resources provided from domesticates. Lim et al. (2021) suggested intron-present genes are more enhancing in transcription and translation than the intron-absent ones. Hence, the introns established in the crop group are plentiful to facilitate that parasitic lifestyle wherein they can reproduce and multiply on ease. In contrast to the number of introns in the *Cardamine* group, merely one or two introns scattered sporadically in the entire rRNA regions but these polymorphic introns carried the putative mobile elements which permit the flexibility of insert themselves into intron-less alleles or novel sites on demands i.e., the aforementioned proto-splice sites located in the rRNA. The flexibility mediated by the spliceosomal activities may be strategic for the survival and reproduction of the pathogens in the face of various environmental circumstances which are not conducive to the hosts (Edgell et al. 2011). For example, two samples (Akiota and Tochigo) from Hiroshima collected from the remotely mountainous areas had different patterns and extra ORFs as opposed to the other four samples (Nara, Nagasaki,

Yamaguchi and Kumamoto) from the *Cardamine* group. Another example of the variable ribosomal intronic regions is the wild weed *Rorippa palustris* **(**marsh yellow cress) (LC716109). The *P. brassicae* population extracted from this cruciferous weed had similar pattern of intronic-absence to many members of the *Cardamine* group. This *P. brassicae* population was also lack of four introns of 18S (intron II and III) and 28S (intron I and II), except for the case of the 18S intron I. The 18S intron I sequence of this population was nearly homologous to the 18S intron I sequences of the *Cardamine* group but this 18S intron I harbored an additional sequence with the intronic ORF (approximately 550 bp) (Supplementary Figure 7). These examples clarify the flexibility and adaptability of intronic regions that are varied by the environmental conditions (within one group of hosts *C. occulta*) or by shifting to a different host, i.e. the weed *R. palustris*. In addition to that, the splicing sites of these spliceosomal introns are not canonical, sites such as TC/GT-TG were found commonly in the *Cardamine* group or even in the crop group. Such non-canonical sites are distinctive to the majority (nearly 99% found commonly introns) having GT at the donor site and AG at the acceptor site (Sibley et al. 2006). They might play certain metabolic functions in *P. brassicae* since the non-canonical splicing sites were known for contributing in diversifying proteomic profiles and involving to exitron, cryptic introns which are responsible for atypical splicing mechanisms (Sibley et al., 2006).

Asides from researching for the metabolic function, the intronic sequences could be a good candidate for polymorphic identification of geographical strains and even linkages to the pathotyping diagnose. Although no mutations have been detected in exonic regions of the LSU, few variations have been detected in the intronic and ITS2 sequences from the isolates AT and e3 (Schwelm et al. 2016). Several populations from around the world were found small differences structured in the secondary ITS2 models (Lam et al. 2022a). In this research, the intronic sequences of the crop group fell in the same narrative that they were

highly identical but few common variations were found in several populations, these variations might be considered as haplotype for certain numbers of Japanese populations.

The pathogenicity of the pathogens *P. brassicae* has been known for its capability of constant change but research to molecular linkages is to no avail due to the complication of pathotyping systems as well as a mixture of several strains in one single gall (Schwelm et al. 2016, Schwelm et al. 2021). Nevertheless, we found few common mutations at both structural sites of the introns or same polymorphic patterns between two groups or even within one group. This may elucidate at least the emergence of novel pathotypes or virulent isolates in the genetic population level. Last but not least, the power of self-splicing introns could also give traces and clues for the sexual reproduction, which is still at debate (Liu et al. 2020), based on the fundamental structures or motifs inherited.

Overall, the role of introns playing life cycle and clubroot disease progression is still indisputable but how its dynamics and evolution characterize in *P. brassicae* is an arena expecting to be discovered.

CHAPTER 5

GENERAL DISCUSSION AND CONCLUSION

The research of clubroot disease occurring on the weed *C. occulta* has actually been conducting for many decades from our group. Due to the limited techniques on molecular basis at that time, our team has not proven these populations are newly discovered in spite of the differences observed on pathological characters from *C. occulta* clubroot. In this research, we proceeded investigating mostly on the genetic aspects of the novel populations but also added additional information of the histophatological descriptions. Our research, as stated repeatedly, was primarily on the cruciferous weed *C. occulta*. This, we hope will get the attention of the clubroot research community from the West whose focus is merely on the Brassica crops. In fact, most current studies on clubroot are from Canada, Europe, the USA…those from the East such as Korea, Japan, China had fewer in number despite these countries especially China, were in the top of brassica crop production (FAO, 2014). In fairness, we found several clubroot articles but most of them published in the native languages of Chinese, Korean or Japanese languages. These articles were in fact providing valuable information for our research. For instance, one Japanses-written article reported mutations in the ITS regions, more interestingly in the 5.8S region. Another Chinese-written one has shown the pathogenicity test (William system) in a clubroot survey, the result also had the pathotype 16 that is similar to our research. The *P. brassicae* extracted from *C. occulta* when tested by the William system, resulted in pathotype 16 (Tanaka and Ito 2013). Lastly, one piece of the information from the reported isolate CD1 (KF129413, retrieved from NCBI Genbank accession) has similar ITS sequences with the Cardamine group. Based on the geographical, cultural and historical relationships between Korea, China and Japan, we can relate to our research.

Although the credibility of these articles should be in doubt but it is inevitable that they broaden the point-of-view of clubroot research and provide many useful pieces of puzzles in the entire picture. Therefore, there should be more collaborations among the clubroot experts from West to East, diversify the aspects of clubroot research. In the end, diversification is still epitome of biology.

For our research, two perspectives should be taken seriously for the *C. occulta* that are the invasiveness and octoploidy genome. The first perspective is obvious that biologist is scared of since invasive species destroy the ecological system. However, for the *C. occulta* that acts as if it were a benevolent host to the *P. brassicae* pathogens. The weeds evolve to fit in the human activities such as agriculture which bring them closer and faster to any cultivated Brassica farms and fields. The latter perspective is the powerful genome of this weed, it is enriched with a large amount of genetic variations. These weeds are still on the course of evolution and so definitely the harbored pathogens. The host-pathogen interactions endown the pathogens varitations, in this case can be interpreted as pathotype generations, population-level diversity. This renders the task of plant breeder, clubroot experts and plant pathologist when investigating on the disease.

Supplementary Tables

Supplementary Table 1 (18S and 28S sequences studied in chapter 4)

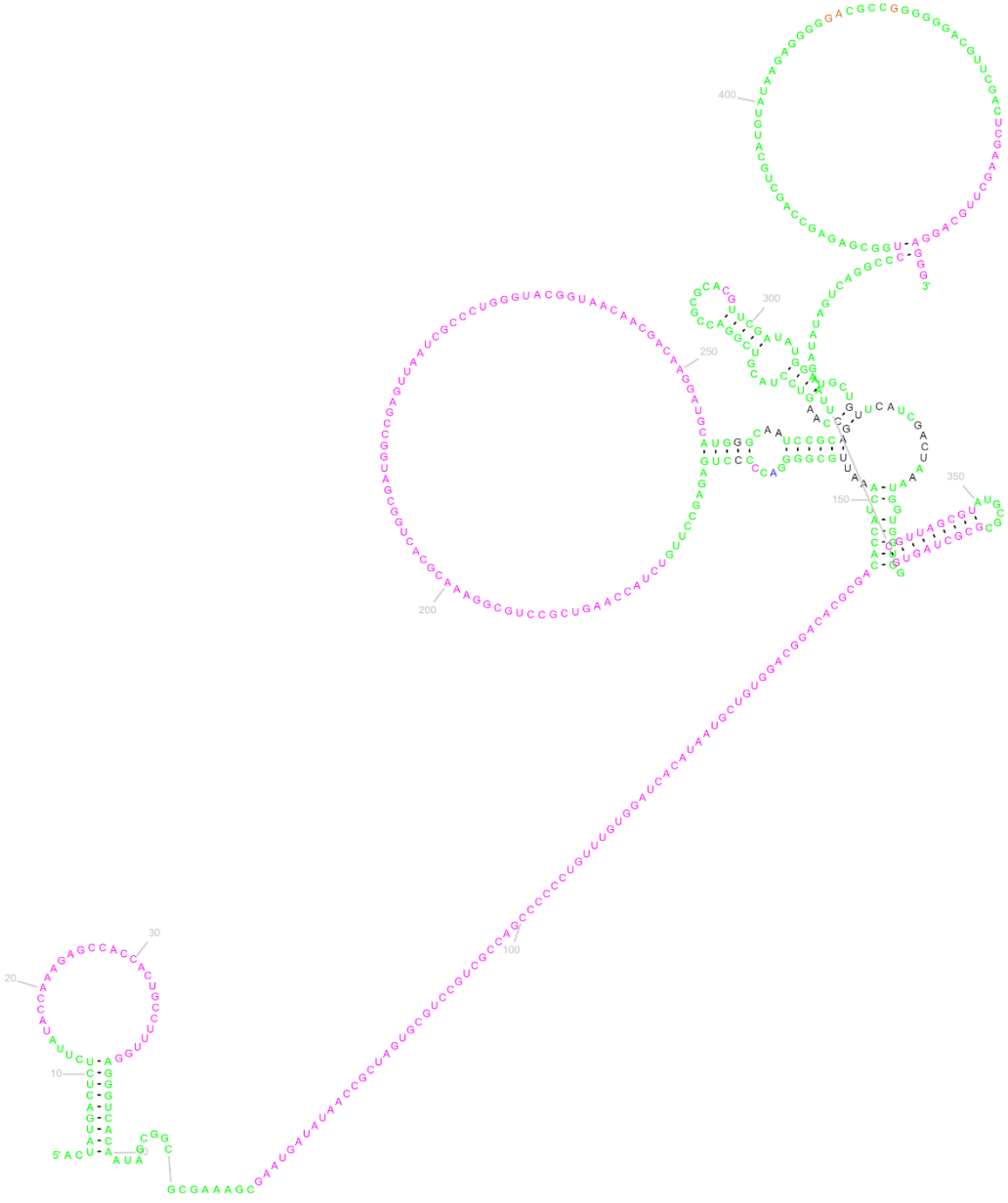
Sequence name	Collection year	Location	Isolate name	Host plant	Accession number
P. brassicae 18S rDNA intron1	2010	Japan:Fukuoka	1Rpalustris	*Rorippa palustris*	LC716109
P. brassicae 18S rDNA intron1	2012	Japan:Nagasaki	1CoNagasaki	*Cardamine occulta*	LC716110
P. brassicae 18S rDNA intron1	1994	Japan:Nara	1CoNara	*Cardamine occulta*	LC716111
P. brassicae 18S rDNA intron1	2013	Japan:Fukuoka	1Fukuoka	*Brassica oleracea var. italica*	LC716112
P. brassicae 18S rDNA intron1	1993	Japan:Ishikawa	1Ishikawa	*Brassica rapa subsp. pekinensis*	LC716113
P. brassicae 18S rDNA intron1	2011	Japan:Saga	1Saga	*Brassica rapa subsp. pekinensis*	LC716114
P. brassicae 18S rDNA intron1	1993	Japan:Aichi	1Aichi	*Brassica rapa subsp. pekinensis*	LC716115
P. brassicae 18S rDNA intron1	1993	Japan:Hokkaido	1Hokkaido	*Brassica rapa subsp. pekinensis*	LC716116
P. brassicae 18S rDNA intron1	1994	Japan:Gunma	1Gunma	*Brassica rapa subsp. pekinensis*	LC716117
P. brassicae 18S rDNA intron1	1994	Japan:Nagasaki	1Nagasaki	*Brassica rapa subsp. pekinensis*	LC716118
P. brassicae 18S rDNA intron1	2014	Japan:Onga	1Onga	*Brassica oleracea var. italica*	LC716119
P. brassicae 18S rDNA intron1	1994	Japan:Ibaraki	1Ibaraki	*Brassica rapa subsp. pekinensis*	LC716120
P. brassicae 18S rDNA intron1	1997	Japan:Osaka	1Osaka	*Brassica rapa subsp. pekinensis*	LC716121
P. brassicae 18S rDNA intron1	2005	Japan:Hagi	1Hagi	*Brassica rapa subsp. pekinensis*	LC716122
P. brassicae 18S rDNA intron1	2014	Japan:Itoshima	1Itoshima	*Brassica oleracea var. italica*	LC716123
P. brassicae 18S rDNA intron2	2005	Japan:Hagi	2Hagi	*Brassica rapa subsp. pekinensis*	LC716124
P. brassicae 18S rDNA intron3	2005	Japan:Hagi	3Hagi	*Brassica rapa subsp. pekinensis*	LC716125
P. brassicae 28S rDNA intron1	1994	Japan:Nara	1CoNara	*Cardamine occulta*	LC716126
P. brassicae 28S rDNA intron1	1994	Japan:Kumamoto	1CoKumamoto	*Cardamine occulta*	LC716127
P. brassicae 28S rDNA intron1	1993	Japan:Yamaguchi	1CoYamaguchi	*Cardamine occulta*	LC716128
P. brassicae 28S rDNA intron1	1994	Japan:Akiota	1CoAkiota	*Cardamine occulta*	LC716129
P. brassicae 28S rDNA intron1	1994	Japan:Togochi	1CoTogochi	*Cardamine occulta*	LC716130
P. brassicae 28S rDNA intron2	2004	Japan:Fukuoka	2Cbpastoris	*Capsella bursa-pastoris*	LC716131
P. brassicae 28S	2005	Japan:Hagi	1Hagi	*Brassica rapa*	LC716132

rDNA intron1				*subsp. pekinensis*	
P. brassicae 28S rDNA intron2	2005	Japan:Hagi	2Hagi	*Brassica rapa subsp. pekinensis*	LC716133
P. brassicae 28S rDNA intron1	2014	Japan:Onga	1Onga	*Brassica oleracea var. italica*	LC716134
P. brassicae 28S rDNA intron2	2014	Japan:Onga	2Onga	*Brassica oleracea var. italica*	LC716135

Supplementary Figures

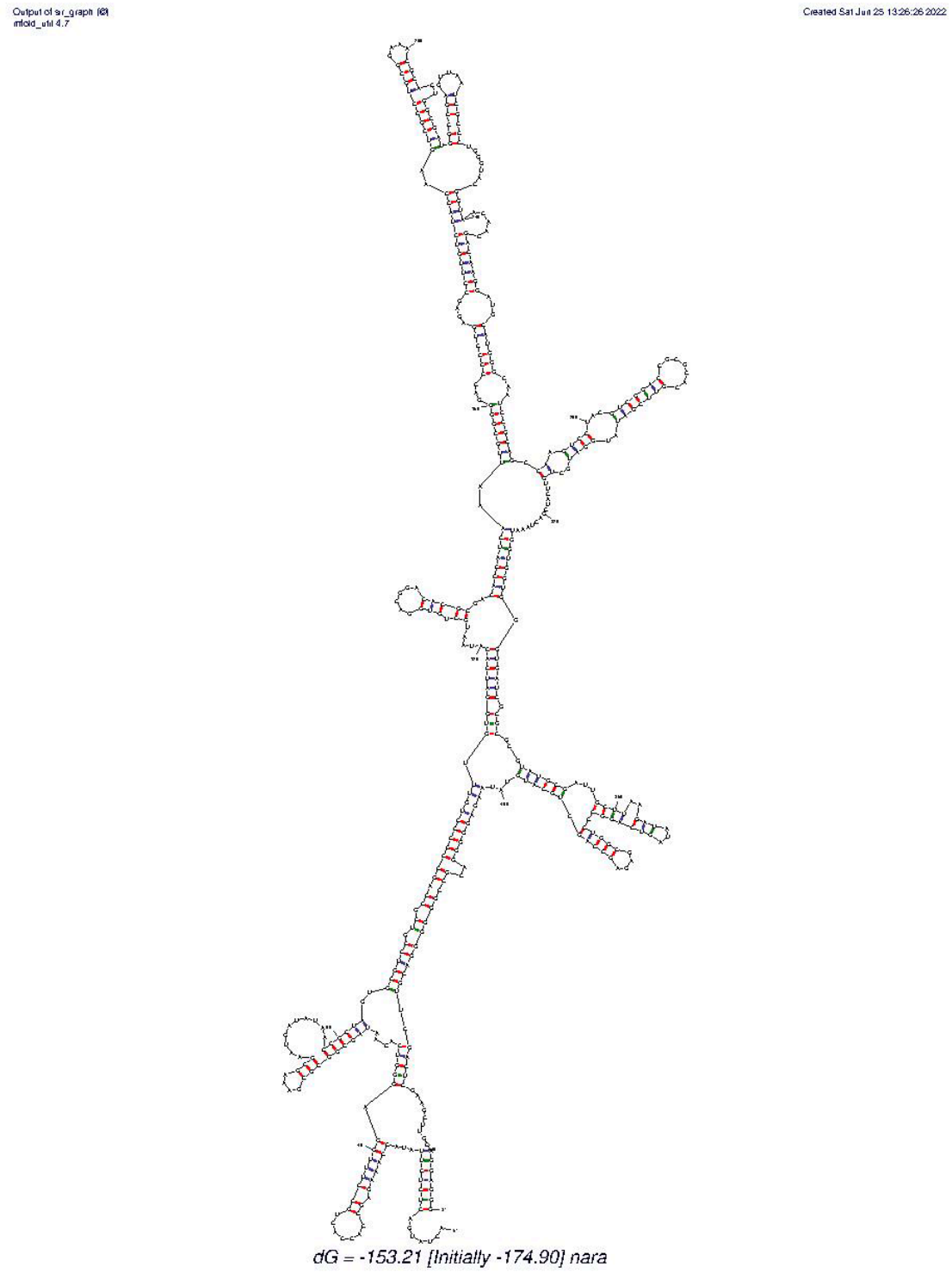
(RNA folded results)

Supplementary Figure 1



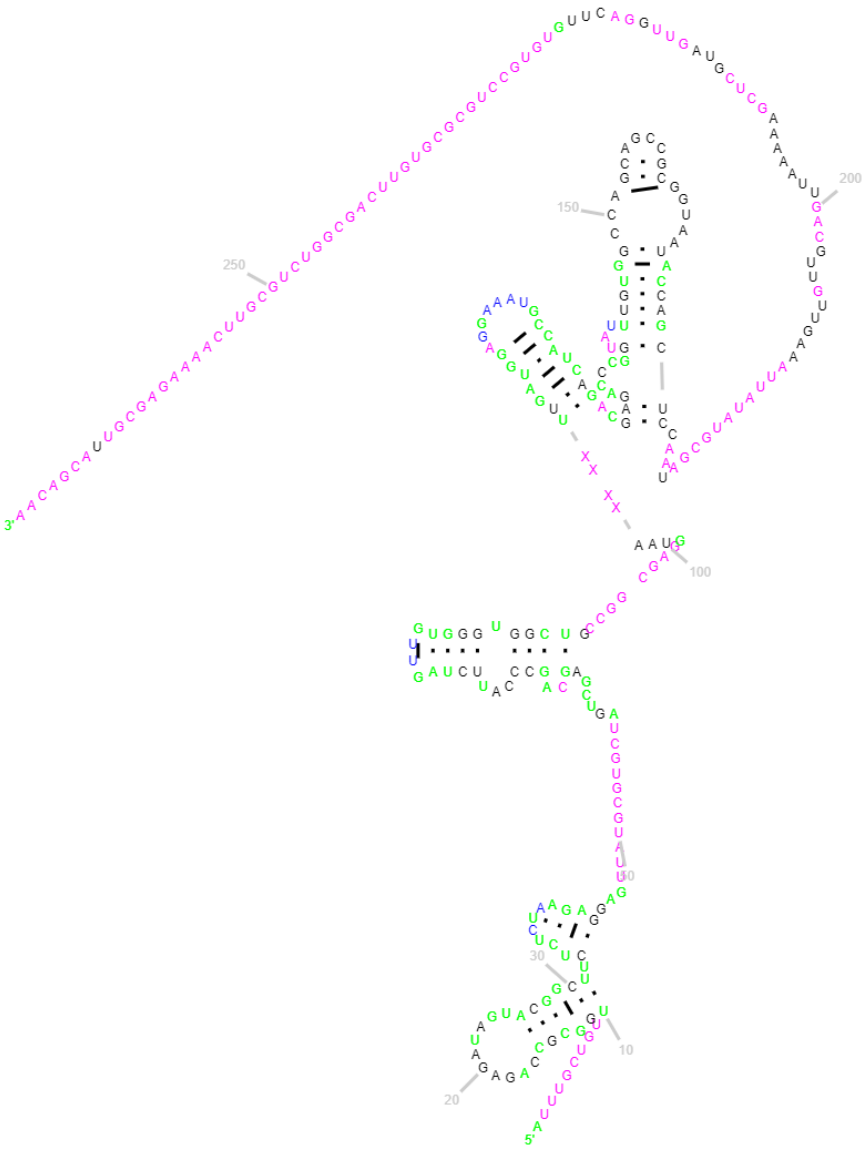
28S1 of the Cardamine group (without the extra sequences) folded by RNA central

Supplementary Figure 2



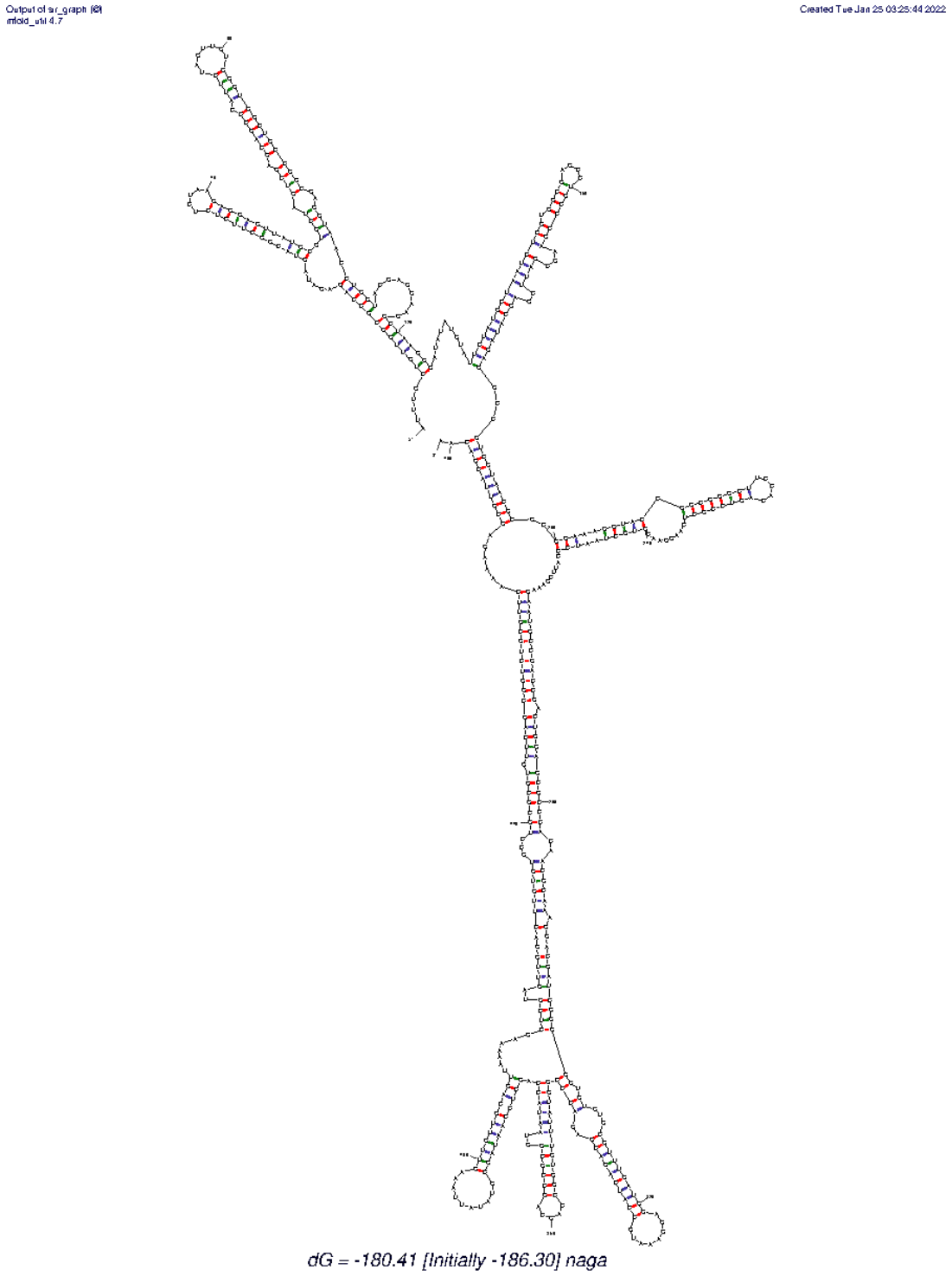
28S1 of the Cardamine group (without the extra sequences) folded by UNAfold with Thermodynamic Entropy (*d*G).

Supplementary Figure 3



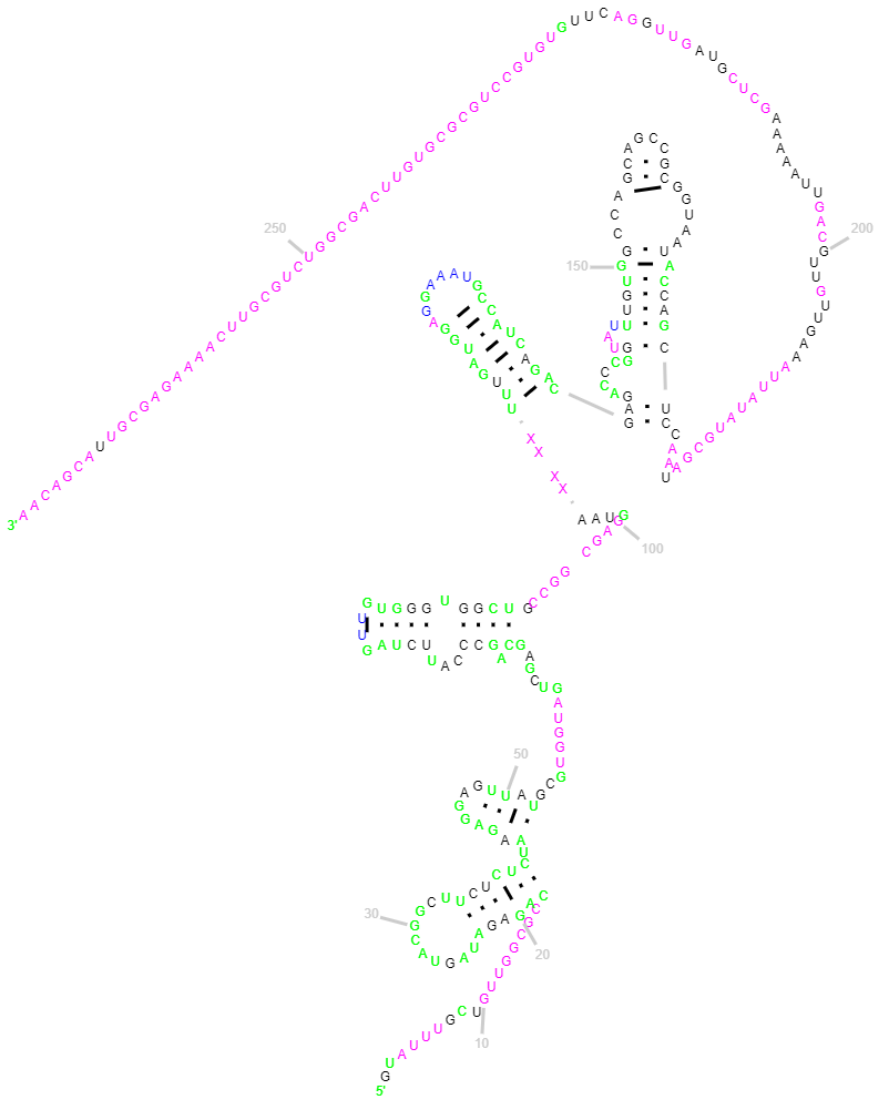
18S1 Nagasaki (Cardamine group) folded by RNA central

Supplementary Figure 4



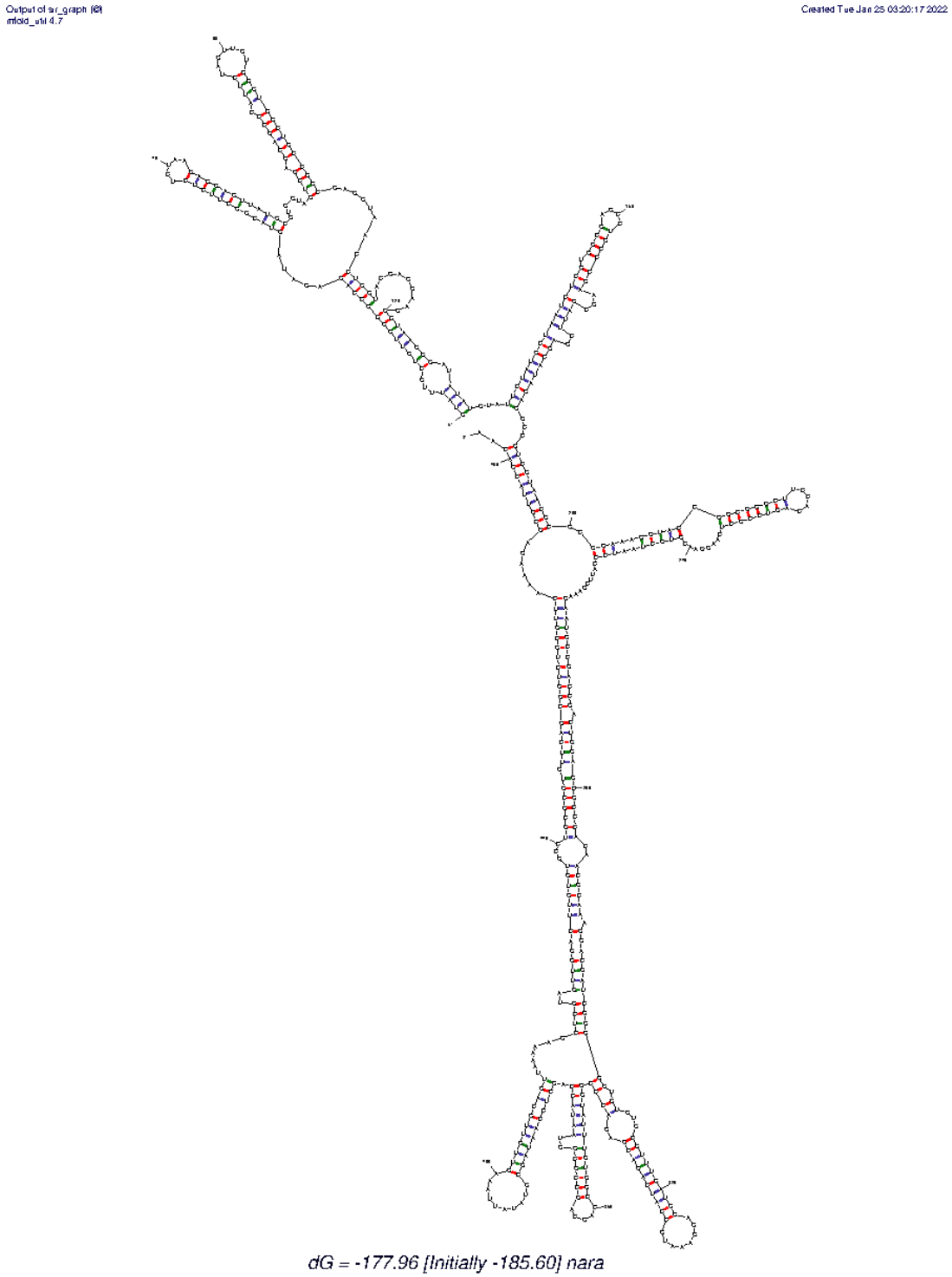
18S1 Nagasaki (Cardamine group) folded by UNAfold with Thermodynamic Entropy (*d*G).

Supplementary Figure 5



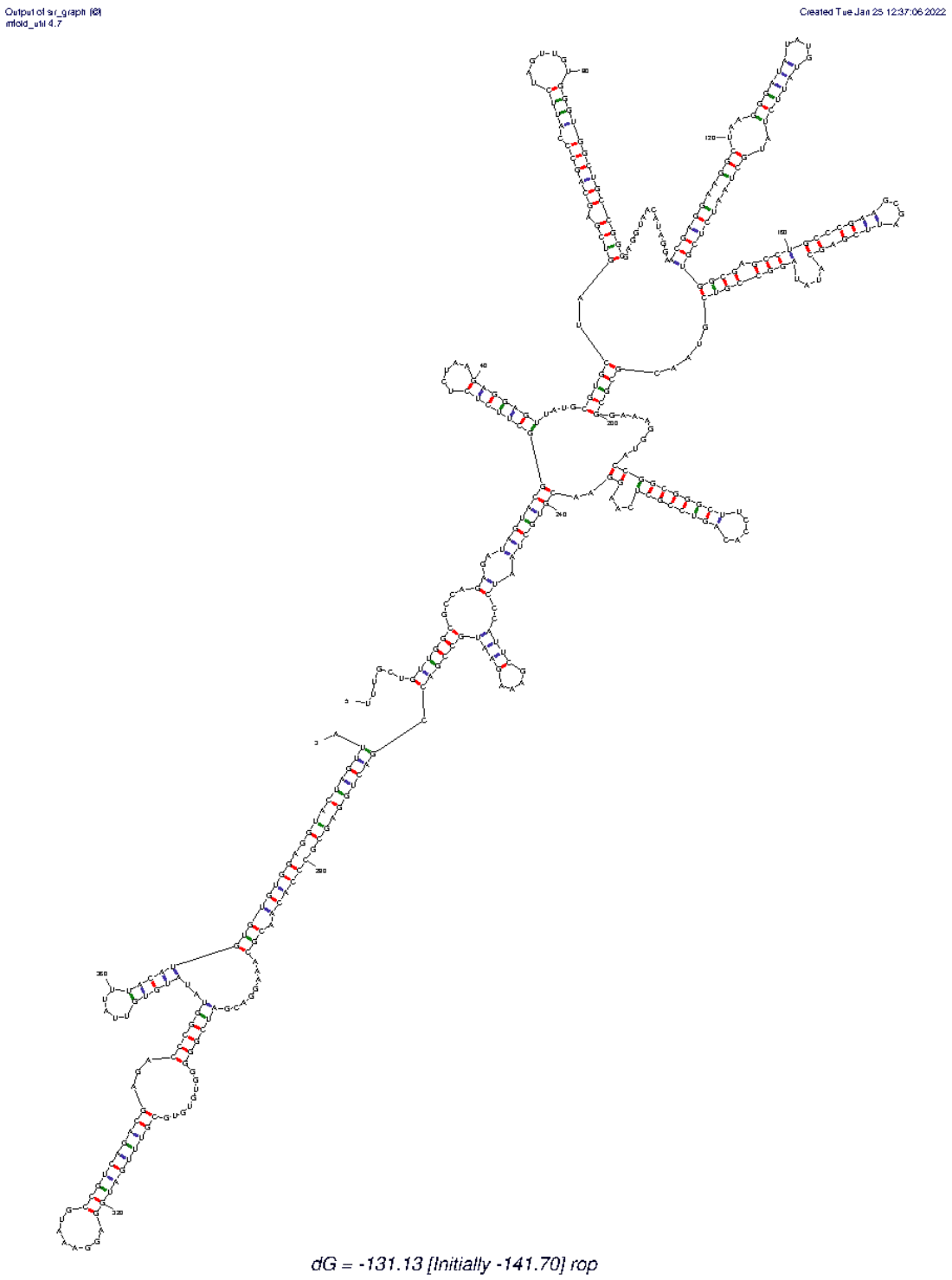
18S1 Nara (Cardamine group) folded by RNA central

Supplementary Figure 6



18S1 Nara (Cardamine group) folded by UNAfold with Thermodynamic Entropy (*d*G).

Supplementary Figure 7



18S1 *Ropripa palutris* without extra sequence folded by UNAfold with Thermodynamic Entropy (*d*G).

Supplementary Figure 8

```
;; Masterfile modified automatically by mfannot version
;; on Wed Jan 19 00:14:52 2022 by user nbeck on host django
;;     - Gene Totals: 1
;;     - List of genes added:
;;       orf120
;;
;; end mfannot
;;


>userdata gc=1
      1  TACCACAGGGGTGGCGGTGGTGGTATATCCCCGCCCGCCCCTGTGAAATCTCAGTGGTAG
     61  CTTTTAAGAGCAGGCTCAAAGTGTCCTG
;      G-orf120 <== end
     89  CTAGTCGGCCATGGTCCGCCTGCTGTCCACCAATGGACCATGGGCACCGGCGACATTGTC
    149  AAATTGCGGGGACATCCTAAAGCCGACTCCACCAAGTCGTCGTGGAAACACGGACGATGG
    209  CCGAGCTAATCGCCCTGGGTATGGTAAAAGCGAG
;; mfannot:      /group=I(derived) (1.42e-08)
    243  tccggatactgaatgggtaatccgcagccaagtcctacggtcgggtcgcccacgcccgat
    303  acggatgcagtccaacgactagatggcaatgggtcgggcatcatactatacgttgcccgg
    363  cttaagatatagtct
;; mfannot:
    378  GGCCCACGGGAAAGCCGTGCCCATCGGATGAGGAGCCCGAGTGGAGGCTCCGGAGCCATG
    438  GGAGGTCGTACCAT
;      G-orf120 <== start
    452  CCAATATATGGGTGTGTGCGACTGGGAGCATGTG
```

MFannot result of 18S1 *Ropripa palustris*

```
;; Masterfile modified automatically by mfannot version
;; on Wed Mar 10 01:48:26 2021 by user nbeck on host django
;;     - Gene Totals: 1
;;     - List of genes added:
;;       orf136
;;
;; end mfannot
;;


>userdata gc=1
      1  ATGACTCTCTTATACCAAAGAGCCACCACTGCCTTTGGAGGGTCACAATAGCGGCGCGAA
     61  AGCGAATGATATAACCGCTAGTGCGTCCGTCGCCAGCCCCCCCTGTTTGTGGATCACATAA
    121  TGCTGTGGACGGACACGCGA
;; mfannot:      /group=IC1 (3.78e-31)
    141  caccatcaaaatgcgggggacaccctgaggagccttgtctaccaagtcgcctgcggaaacgc
    201  actggcgatggccggagttaatcgccctgggtacggtaacaacgacaaggatgcatgggca
    261  atccgcagccaagtcctacgtcggaccgcgcacgttcgatatggatgctgttcatcgact
    321  aaatggtggtgggtgatcgcgcgcgtatgcgattgcttaagatatagtcaggc
;; mfannot:
    374  CCTGGCGAGAGCCAGCTGCATGTATAAGAGGGGGACGCCGGGGGGACGTTCGACTCGAAG
    434  CTTGCAGGAGGGCGCTGTTTACTTGTCGTAGCAGCGTCTTTGGAAGGCGCTTCAGTATTT
    494  AGGAATGAAGCCGTCAGTTAAGCGCCAGAATAAAAGATGTTGGGGTCGTTGAACATGTCA
    554  GTTGGCAGAATGTGGTCCAACTCATCGTCTCCTAAGGTCGAAGGACATGTTCCGATGTCG
    614  TCGTCCTGAATGGGATTGCCGATTACAGTACCGTTGAGCGCATTGTTCAGGATATCAGTG
    674  GCTTTGTTGCCGATGTTGTTCTTGATCATCCGCACGCCACGTCCTGCACGAATGCATCGT
    734  GGTTTGTGGACGCGCACGTAGCTTTGCCCAGGCAGAAGTTGCGGCTCTTG
;      G-orf136 <== end
    784  TTAACGTCGAGTGATTCGTACACGATGTGGACGAGGCGAACACAACAACGGTCGTGATGG
    844  CATAGATGAGGTGAGGCTACTGGATGATCGAACAGGATCAGTTGGGACGCCAGAGTTAAC
    904  ATACTGCTGGAGTACTAACGGCACGTTCGTCGCCATGACTCTGACCCGCTGTGATAGACC
    964  AAGCAGTTTAGTGATATCGAACTGTGGGCGACCATTGCTGTTGTTGGAGAAACCCACTAT
   1024  GGGATTATCCACATAGAGGCATTCGATTGCTAAATAAGTGGTGGGCTCACCAGAAGAGGA
   1084  TATGCTCTCTGGGACCCAGTAGTGCTTGGTCACAGCACTTTGGAGCCTCGCTGTAAACGT
   1144  CCGCCAAGTATAGTTCGCCTGGCGACGACGACGTTGCACGAGGTCCGACAT
;      G-orf136 <== start
   1195  ACAAACGCACCCAACGCAGCGCTTTCTGACGACGGACTCCGCGACACTGATCCAACGCGC
   1255  ACACACACTG
```

MFannot result of 28S1 of Akiota and Togochi

REFERENCES

Alvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. Mol Phylogenet Evol 23:417-434

Al-Shehbaz IA, Beilstein MA, Kellogg EA (2006) Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. Plant Syst Evol 259:89–120

Alizon S, Hurford A, Mideo N, Van Baalen M (2009) Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. J Evol Biol 22(2):245-59

Anon (2000) (The Arabidopsis genome initiative partnership). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nat 408:796–815

Agrios GN (2005) Plant Pathol. 5th Edition, Elsevier Academic Press, Amsterdam

Archibald JM and Keeling PJ (2004) Actin and ubiquitin protein sequences support a cercozoan/foraminiferan ancestry for the Plasmodiophorid Plant Pathogens. J Eukaryot Microbiol 51: 113-118

Ardenghi NMG, Mossini S (2014) *Cardamine flexuosa subsp. debilis* O. E. Schulz. In: von Raab-Straube E, Raus T (Eds) Euro+Med-Checklist Notulae, 3. Willdenowia 44: 292.

Bass D, Moreira D, Lopez-Garcia P, Polet S, Chao EE, Von der Heyden S, Pawlowski J, CavalierSmith T (2005) Polyubiquitin insertions and the phylogeny of Cercozoa and Rhizaria. Protist 156:149–161

Belshaw R, Bensasson D (2006) The rise and falls of introns. Heredity 96:208–213

Bi K, He Z, Gao Z (2016) Integrated omics study of lipid droplets from *Plasmodiophora brassicae*. Sci Rep 6 (36965)

Bleeker W, Klausmeyer S, Peintinger M, Dienst M (2008) DNA sequences identify invasive alien Cardamine at Lake Constance. Biolo Conserv 141: 692–698

Bulman S, Ridgway HJ, Eady C, Conner AJ (2007) Intron-rich gene structure in the intracellular plant parasite *Plasmodiophora brassicae*. Protist 158:423-33

Bulman S, Richter F, Marschollek S, Benade F, Jülke S, Ludwig-Müller J (2019) *Arabidopsis thaliana* expressing PbBSMT, a gene encoding a SABATH-type methyltransferase from the plant pathogenic protest *Plasmodiophora brassicae*, show leaf chlorosis and altered host susceptibility. Plant Biol J 21: 120-130

Buczacki ST, Toxopeus H, Mattusch P, Johnston TD, Dixon GR et al. (1975) Study of physiological specialization in *Plasmodiophora brassicae*: proposals for attempted rationalization through an international approach. Trans Br Mycol Soc 65:295–303

Burki F, Keeling PJ (2014) Rhizaria. Curr Biol 24:R103–R107

Cavalier-Smith T (1993) Kingdom protozoa and its 18 phyla. Microbiol Rev 57:953–994

Cavalier-Smith T (2002) The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int J Syst Evol Microbiol 52:297–354

Coleman AW (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends Genet 7:370-375

Crete R (1981) Worldwide importance of clubroot. Clubroot Newslett 11:6–7

Crow KD, Wagner GP (2006) What Is the Role of Genome Duplication in the Evolution of Complexity and Diversity. Mol Biol Evol 23:887–892

Daval S, Belcour A, Gazengel K, Legrand L, Gouzy J, Cottret L, Lebreton L, Aigu Y, Mougel C, Manzanares-Dauleux MJ (2019) Computational analysis of the *Plasmodiophora brassicae* genome: mitochondrial sequence description and metabolic pathway database design. Genomics 111(6):1629-1640

Diederichsen E, Frauen M, Linders E, Hatakeyama K, Hirai M (2009) Status and perspectives of clubroot resistance breeding in crucifer crops. J Plant Growth Regul 28:265–281

Dixon GR (1991) Primary and secondary stages of *Plasmodiophora brassicae* (clubroot) affected by metallic cations and pH. Deve Agr Magn Fore Eco 23:381-386

Dixon GR (2009a) *Plasmodiophora brassicae* in its environment. J Plant Growth Regul 28:212–228

Dixon GR (2009b) Husbandry the sustainable means of controlling soil borne pathogens: asynoptic review. Acta Hortic 817:233–242

Dixon GR (2009c) The occurrence and impact of *Plasmodiophora brassicae* and clubroot disease. JPlant Growth Regul 28:194–202

Dixon GR (2002) Interactions of soil nutrient environment, pathogenesis and host resistance. Plant Prot Sci 38(1):87–94

de Souza SJ, Long M, Klein RJ, Roy S, Lin S (1998) Toward a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. Proc Natl Acad Sci USA 95: 5094–5099

Donald C, Porter IJ (2004) A sand solution culture technique used to observe the effect of calcium and pH on root hair and cortical stages of infection by *Plasmodiophora brassicae*. Aus Plant Pathol 33(4):585–589

Ebert D, Fields PD (2020) Host-parasite co-evolution and its genomic signature. Nat Rev Genet 21:754-768

Edgell DR., Chalamcharla, VR, Belfort M (2011) Learning to live together: mutualism between self-splicing introns and their hosts. BMC Biol 9:22

Esser HJ (2020) *Cardamine occulta* (Brassicaceae), a new name for the Flora of Thailand. Thai Fore Bullet (Botany*) 48*(2):187–189.

Feng J, Jiang J, Feindel D, Strelkov SE, Hwang SF (2016) The gene Cr811 is present exclusively in pathotype 5 and new emerged pathotypes of the clubroot pathogen *Plasmodiophora brassicae*. Eur J Plant Pathol 145:615–620

Flipphi M, Fekete E, Ag N, Scazzocchio C, Karaffa L (2013) Spliceosome twin introns in fungal nuclear transcripts. Fungal Genet Biol 57:48-57

Flipphi M, Ág N, Karaffa L, Kavalecz N, Cerqueira G, Scazzocchio C, Fekete E (2017) Emergence and loss of spliceosomal twin introns. Fungal Biol Biotechnol 4:7

Flint-Garcia SA (2013) Genetics and consequences of crop domestication. J Agric Food Chem 61:8267-76

Gahatraj S, Shrestha SM, Devkota TR, Rai HH (2019) A review on clubroot of crucifers: symptoms, life-cycle of pathogen, factors affecting severity, and management strategies. Arch Agricul Environ Sci 4 342–349

Gibbs JG (1932) Weed host plants of club-root in New Zealand. NZ J Agric 44:273–276

Goeche F, Wiese J, Núnez A, Labes A, Imhoff JF, Neuhauser S (2012) A novel phytomyxean parasite associated with galls on the bull-kelp *Durvillaea antarctica* (Chamisso) Hariot. PLoS ONE 7:e45358

Govaerts R, 1999: World Checklist of Seed Plants 3(1, 2a & 2b). MIM, Deurne, p. 1–1532

Guillén D, Sánchez S, and Rodríguez-Sanoja R (2010) Carbohydrate-binding domains: Multiplicity of biological roles. Appl Microbiol Biotechnol 85:1241-1249

Heenan PB (2017) A taxonomic revision of Cardamine L. (Brassicaceae) in New Zealand. Phytotaxa 330(1): 001–154.

Hruševar D, Mesaroš J, Vladović D, Vucić A, Belamarić I, Surać L, Mitić B (2021) *Cardamine occulta* Hornem. –a new concealed alien plant in the flora of Croatia. Natura Croatica: Periodicum Musei Historiae Naturalis Croatici 30(1):207-215.

Hwang SF, Ahmed HU, Zhou Q, Strelkov SE, Gossen BD, Peng G, Turnbull GD (2011) Influence of cultivar resistance and inoculum density on root hair infection of canola (*Brassica napus*) by *Plasmodiophora brassicae*. Plant Pathol 60: 820-829

Hwang SF, Strelkov SE, Feng J, Gossen BD, Howard RJ (2012) *Plasmodiophora brassicae*: a review of an emerging pathogen of the Canadian canola (*Brassica napus*) crop. Mol Plant Pathol 3(2):105-13

Ito S, Maehara T, Tanaka S, Kameya-Iwaki M, Yano S, Kishi F (1997) Cloning of a single-copy DNA sequence unique to *Plasmodiophora brassicae*. Physiol Mol Plant Pathol 50:289-300

Jiang X, Su Y, Wang M (2022) Mapping of a novel clubroot disease resistance locus in *Brassica napus* and related functional identification. Front Plant Sci 28; 13:1014376.

Jing W, Yun H, Xiaoling H, Yingze N, Xiaolan L, Yong L (2008) Study of symptoms, yield loss of clubroot and modality of *Plasmodiophora brassicae* in rape. Chin J Oil Seed Sci 30(1):112–115

Jörg S, Tobias M, Marco A, Philipp NS, Thomas D, Matthias W (2006) The internal transcribed spacer 2 database—a web server for (not only) low level phylogenetic analyses. Nucleic Acids Res 34:704-707

Kagale S, Robinson SJ, Nixon J, Xiao R, Huebert T, Condie J, Kessler D, Clarke WE, Edger PP, Links MG, Sharpe AG, Parkin IA (2014) Polyploid evolution of the Brassicaceae during the Cenozoic era. Plant Cell 26(7):2777-91

Kageyama K, Asano T (2009) Life cycle of *Plasmodiophora brassicae*. J Plant Growth Regul 28:203–211

Keller A, Förster F, Müller T, Dandekar T, Schultz J, Wolf M (2010) Including RNA secondary structures improves accuracy and robustness in reconstruction of

phylogenetic trees. Biol Direct 5: 4

Kim H, Jo EJ, Yong HC, Jang KS, Choi GJ (2016) Pathotype classification of *Plasmodiophora brassicae* isolates using clubroot-resistant cultivars of Chinese cabbage. Plant Pathol J 32:423–430

Koonin EV (2006) The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? Biol Direct 1: 22

Kudoh H (2017) Biology of the weedy species of the genus *Cardamine* [Brassicaceae] in Japan. J Weed Sci and Tech 62:175-18

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547-1549

Laila R, Robin AH, Yang K, Choi GJ, Park JI, Nou IS (2017) Detection of Ribosomal DNA Sequence Polymorphisms in the Protist *Plasmodiophora brassicae* for the Identification of Geographical Isolates. Int J Mol Sci 18:84.

Lam ATP, Sasaki K, Yanagi Y, Tanaka S, Ito S (2022a) Genotypic characterization of *Plasmodiophora brassicae* in the paddy-field weed *Cardamine occulta* and symptomology reveal a distinct pathogen population in Japan. J Gen Plant Pathol

Lam ATP, Sasaki K, Yanagi Y, Tanaka S, Ito S (2022b) Intraspecific Variability and Distribution Difference within the Ribosomal Introns of the Discrete *Plasmodiophora brassicae* Group in Japan: A Case Study for Complex Dynamics of Intron Evolution. Agronomy 12(9):2154.

Leostrin AV, Mayorov SR (2019) Current state and distribution of alien weedy *Cardamine occulta* Hornem. (Brassicaceae) in European Russia. Russ Journ of Biolo Invas 10: 236–245

Lihová J, Marhold K, Kudoh H Koch, MA (2006) Worldwide phylogeny and biogeography of *Cardamine flexuosa* (Brassicaceae) and its relatives. Am J Bot 93:1206-1221

Lim CS, Weinstein BN, Roy SW, Brown CM (2021) Analysis of Fungal Genomes Reveals Commonalities of Intron Gain or Loss and Functions in Intron-Poor Species. Mol Biol Evol 38:4166-4186

Liu SL, Adams KL (2010) Dramatic change in function and expression pattern of a gene duplicated by polyploidy created a paternal effect gene in the Brassicaceae. Genome Biol Evol 27:2817–2828

Liu L, Qin L, Zhou Z, Hendriks W, Liu S, Wei Y (2020) Refining the Life Cycle of *Plasmodiophora brassicae*. Phytopathology 110:1704–1712.

Long M, Rosenberg C (2000) Testing the "proto-splice sites" model of intron origin: evidence from analysis of intron phase correlations. Mol Biol Evol 17:1789-96

Ludwig-Müller J, Prinsen E, Rolfe SA, Scholes JD (2009) Metabolism and plant hormone action during clubroot disease. J Plant Growth Reg. 28:229–244

Mandáková T, Zozomová-Lihová J, Kudoh H, Zhao YP, Lysak MA, Marhold K (2019) The story of promiscuous crucifers: origin and genome evolution of an invasive species, *Cardamine occulta* (Brassicaceae), and its relatives. Ann Bot 124:209–220

Mansanet-Salvador CJ, Ferrer-Gallego PP, Ferrando I Laguna E (2015) Notas sobre el complejo taxonómico *Cardamine flexuosa* With. (Cruciferae) y su presencia en la Comunidad Valenciana. Flora Montiberica 59: 72–82

Marhold K, Šlenker M, Kudoh H, Zozomová-Lihová J (2016) *Cardamine occulta*, the correct species name for invasive Asian plants previously classified as *C. flexuosa*, and its occurrence in Europe. PhytoKeys 5:57-72

Märkle H, John S, Cornille A, Fields PD, Tellier A (2021) Novel genomic approaches to study antagonistic coevolution between hosts and parasites. Mol Ecol 30:3660– 3676

McDonald BA, Linde C (2002) Pathogen population genetics, evolutionary potential, and durable resistance. Annu Rev Phytopathol 40:349-379

Morita S, Azuma M, Aoba T, Satou H, Narisawa K, Hashiba T (2003) Induced systemic resistance of Chinese cabbage to bacterial leaf spot and Alternaria leaf spot by the root endophytic fungus, *Heteroconium chaetospira*. J Gen Plant Pathol 69:71–75

Nikolaev SI, Berney C, Fahrni JF, Bolivar I, Polet S, Mylnikov AP, Aleshin VV, Petrov NB, Pawlowski J (2004) The twilight of heliozoa and rise of Rhizaria, an emerging super group of amoeboid eukaryotes. Proc Natl Acad Sci USA 101:8066–8071

Niwa R, Kawahara A, Murakami H, Tanaka S, Ezawa T (2011) Complete structure of nuclear rDNA of the obligate plant parasite *Plasmodiophora brassicae*: intraspecific polymorphisms in the exon and group I intron of the large subunit rDNA. Protist 162(3):423-34.

Nowicki B (1973) Host range of *Plasmodiophora brassicae* Wor. (in Polish). Acta Agrobot 26:53–61

Pliszko A. (2020) First record of Asian Cardamine occulta Hornem. (Brassicaceae) in Poland. BioInvasions Records 9(3): 655-659.

Prahl RE, Khan S, Deo RC (2021) The role of internal transcribed spacer 2 secondary structures in classifying mycoparasitic Ampelomyces. PLoS One 16(6):e0253772.

Schwelm A, Fogelqvist J, Knaust A, Jülke S, Lilja T, Bonillarosso G, Karlsson M, Shevchenko A, Dhandapani V, Choi SR, Kim HG, Park JY, Lim YP, Ludwig-Muller J, Dixelius C (2015) The *Plasmodiophora brassicae* genome reveals insights in its life cycle and ancestry of chitin synthases. Sci Rep 5:11153

Saharan G, Mehta N, Meena PD (2021) Clubroot Disease of Crucifers: Biology, Ecology and Disease Management.

Schuller A and Ludwig-Müller J (2016) Histological methods to detect the clubroot pathogen *Plasmodiophora brassicae* during its complex life cycle. Plant Pathol, 65: 1223-1237.

Schwelm A, Fogelqvist J, Knaust A *et al.* (2015) The *Plasmodiophora brassicae* genome reveals insights in its life cycle and ancestry of chitin synthases. *Sci Rep* **5**, 11153

Schwelm A, Berney C, Dixelius C, Bass D, Neuhauser S (2016) The Large Subunit rDNA sequence of *Plasmodiophora brassicae* does not contain intra-species polymorphism. Protist 16:544-554

Schwelm A, Neuhauser S (2017) Letter to the Editor: "Detection of Ribosomal DNA Sequence Polymorphisms in the Protist *Plasmodiophora brassicae* for the Identification of Geographical Isolates". Int Jour Mol Sci 18(7):1454

Schwelm A, Ludwig-Müller J (2021) Molecular Pathotyping of *Plasmodiophora brassicae*-Genomes, Marker Genes, and Obstacles. Pathogens 10:259.

Sibley CR, Blazquez L, Ule J (2016) Lessons from non-canonical splicing. Nat Rev Genet 17:407–421

Siemens J, Bulman S, Rehn F, Sundelin T (2009) Molecular biology of *Plasmodiophora brassicae*. J Plant Growth Regul 28:245–251

Šlenker M, Zozomová-Lihová J, Mandáková T, Kudoh H, Zhao Y, Soejima A, Yahara T, Skokanová K, Španiel S, Marhold K (2018) Morphology and genome

size of the widespread weed *Cardamine occulta*: how it differs from cleistogamic *C. kokaiensis* and other closely related taxa in Europe and Asia. Botanical Journ of the Linnean Soc 187(3): 456–482.

Šlenker M, Zozomová-Lihová J, Marhold K (2019) *Cardamine occulta*-inconspicuous neophyte in Slovakia. Bulletin Slovenskej Botanickej Spoločnosti 41(1): 13–23

Shimizu-Inatsugi R, Terada A, Hirose K, Kudoh H, Sese J, Shimizu KK (2017) Plant adaptive radiation mediated by polyploid plasticity in transcriptomes. Mol Ecol 26(1):193-207.

Stjelja S, Fogelqvist J, Tellgren-Roth C (2019) The architecture of the *Plasmodiophora brassicae* nuclear and mitochondrial genomes. Sci Rep 9:15753

Struck C, Rüsch S, Strehlow B (2022) Control Strategies of Clubroot Disease caused by *Plasmodiophora brassicae*. Microorganisms 10(3):620

Rakow G (2004) Species origin and economic importance of Brassica. In: Pua EC, Douglas CJ (eds) Biotechnology in agriculture and forestry, vol 54, 1st edn. Springer, Berlin, pp 3–11

Ren L, Xu L, Liu F, Chen K, Sun C, Li J, Fang X (2016) Host Range of Plasmodiophora brassicae on Cruciferous Crops and Weeds in China. Plant Dis. 100(5):933-939

Riascos D, Ortiz E, Quintero D, Montoya L, Hoyos-Carvajal L (2011) Histopathological and morphological alterations caused by *Plasmodiophora brassicae* in *Brassica oleracea* L. Agron Colomb 29:57–61

Rogers SO (2019) Integrated evolution of ribosomal RNAs, introns, and intron nurseries. Genetica 147:103–119

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. Mol Biol Evol 34: 3299-3302

Rogozin IB, Carmel L, Csuro M (2012) Origin and evolution of spliceosomal introns. Biol Direct 7:11

Roy SW, Gilbert W (2005) The pattern of intron loss. Proc Natl Acad Sci USA 102:713-8

Roy SW, Hartl DL (2006) Very little intron loss/gain in Plasmodium: intron loss/gain mutation rates and intron number. Genome Res 16:750-756

Roy SW, Penny D (2007) Widespread intron loss suggests retrotransposon activity in ancient apicomplexans. Mol Biol Evol 24:1926-33

Rolfe SA, Strelkov SE, Links MG (2016) The compact genome of the plant pathogen *Plasmodiophora brassicae* is adapted to intracellular interactions with host *Brassica spp*. BMC Genomics 17:272

Rutland CA, Hall ND, McElroy JS (2021) The Impact of Polyploidization on the Evolution of Weed Species: Historical Understanding and Current Limitations. Front Agron 3:626454

Tanaka S, Negoro M, Ota T, Katumoto K, Nishi Y (1990) Clubroot of spring Chinese cabbage in Nagasaki Prefecture, Kyushu (in Japanese with English summary). Bull Fac Agric Yamaguchi Univ 38:33–45

Tanaka S, Ito S, Kameya-Iwaki M, Katumoto K, Nishi Y (1993) Occurrence and distribution ofclubroot disease on two cruciferous weeds, *Cardamine flexuosa* and *C. scutata*, in Japan. Trans Mycol Soc 34:381–388

Tanaka S, Murai K, Ito S, Katumoto K, Nishi Y (1994) The occurrence clubroot disease of wasabi (*Eutrema wasabi* Maxim.) and its possible source of infection (in Japanese with English summary). Ann Phytopathol Soc 60:257–259

Tanaka S, Mizui Y, Terasaki H, Ito S (2006) Distribution of clubroot disease of a cruciferous weed, Cardamine flexuosa, in major isolated islands, Hokkaido and Okinawa in Japan. Mycoscience 47(2):72–77

Tanaka S, Ito S (2013) Pathogenic and genetic diversity in *Plasmodiophora brassicae* (clubroot) from Japan. J Gen Plant Pathol 79(5):297–306

Wang Y, Wu F, Bai J, He Y (2014) BrpSPL9 (*Brassica rapa ssp. pekinensis* SPL9) controls the earliness of heading time in Chinese cabbage. Plant Biotechnol J 12:312–321

Williams PH (1966) A system for the determination of races of *Plasmodiophora brassicae* that infect cabbage and rutabaga. Phytopathology 56:624–626

Yin LF, Hu MJ, Wang F, Kuang H, Zhang Y (2012) Frequent Gain and Loss of Introns in Fungal Cytochrome b Genes. PLOS ONE 7:e49096

Yatsu Y, Kachi N, Kudoh H (2003) Ecological distribution and phenology of an invasive species, *Cardamine hirsuta* L., and its native counterpart, *Cardamine flexuosa* With., in central Japan. Plant Spec Biol 18: 35-42

Yi Z, Huang L, Yang R, Lin X, Song W (2016) Actin evolution in ciliates (Protist, Alveolata) is characterized by high diversity and three duplication events. Mol Phylogenet Evol 96:45-54.

Zhang H, Feng J, Zhang S, Zhang S, Li F, Strelkov SE, Sun R, Hwang SF (2015) Resistance to *Plasmodiophora brassicae* in *Brassica rapa* and *Brassica juncea* genotypes from China. Plant Dis 99:776–779

Zhang J, Zhang YZ, Jiang J, Duan CG (2020 The Crosstalk Between Epigenetic Mechanisms and Alternative RNA Processing Regulation. Front Genet 20(11):998

Zheng J, Wang X, Xiao Y, Wei S, Wang D, Huang Y (2018/19) Specific genes identified in pathotype 4 of the clubroot pathogen *Plasmodiophora brassicae*. Plant Dis 103(3):1–9

Zhou Q, Hwang SF, Strelkov SE, Fredua-Agyeman R, Manolii VP (2018) A Molecular Marker for the Specific Detection of New Pathotype 5-like Strains of *Plasmodiophora brassicae* in Canola. Plant Pathol 67:1582–1588

SUMMARY

Clubroot disease causes great damages to the *Brassica* crops. The causal pathogen is *Plasmodiphora brassicae*, a telluric intracellular obligate protist whose life cycle and pathogenesis remain largely unknown. The mainstream research on clubroot merely focuses on the *Brassica* crops; however, our clubroot research on the cruciferous weed *Cardamine occulta*, commonly grown in rice paddy fields in Japan, has revealed novel populations *P. brassicae* (a.k.a *Cardamine* group). The wild weed *C. occulta*, is an overwinter perennial crucifer whose genome and life cycle have coevolve with human activities, e.g. agricultural activities. This weed possesses an octoploidy genome that enriched in genetic variations and mutations. These advantageous elements help this weed adapt well into the new ecological niches, especially for anthropogenic habitats such as rice fields, nursing plant houses, construction sites, etc…. Simultaneously, the genetic variations of the weeds, via the host-pathogen interactions, also enhance the genetic diversification of the *P. brassicae* populations. In the present study, the coevolution relationship was approached basically under the scope of the pathogen *P. brassicae* which was molecularly investigated in addition to pathological traits description.

The first and foremost approach is those pathological traits including morphological and histopathological description were demonstrated in comparison with those occurred on the *Brassica* crops. In that sense, few distinctive characters were observed only in the *Cardamine* group. For the morphological characters, large galls were observed on the stem (stalk) of the weed as well as above-ground symptoms cannot be observed despite heavily affected. For microscopic examination, plasmodia were detected in vascular bundles and pith tissues and even in the foliage structure such as petioles. These characters of symptomology are the signals for the genetic investigations in order to prove the *P. brassicae* infected *C. occulta* displays as novel populations.

The second approach is molecular examinations to corroborate that pathological characterization by proving the genetic identity of the novel populations. *Actin* genes and ITS regions were selected to prove the genetic distinction of the *Cardamine* group from the cosmopolitan strains (*Brassica* group).

The *actin* genes investigation, known for significant role in several cellular functions and pathogenesis in *P. brassicae*, revealed the *Cardamine* group accumulated seven point mutations (three silent and four missense) in comparison with the *actin* genes of the cosmopolitan strains.

For ITS regions, popularly known in phylogenetics, the results demonstrated in both ITS1 and ITS2 regions acquired large amount of mutations. A greater number of variations and mutations (single-nucleotide polymorphisms, oligonucleotide polymorphisms, and insertions/deletion) were found in the ITS2 than those in ITS1 and these mutations acquired were greater than any cosmopolitan *P. brassicae* infected on the *Brassica* crops. Concatenated phylogenetic analysis conducted on both *actin* and ITS sequences has confirmed the separated status of the *Cardamine* groups from the global strains. Since the ITS2 sequences were enriched with large number of mutations, further structurally analyzed was deployed to reinforce the phylogenetic assessment. The secondary structure analyses of ITS2 obtained by the *Cardamine* group revealed it was greatly distanced from the ITS2 secondary structure of the *Brassica* group. The *Cardamine* group was different structured in base ring, size of four helices comprised of bulges and loops (based on the nucleotides bonds) as opposed to fundamental structure formed by the cosmopolitan strains.

Further molecular assessment was deployed based on the extended intergenic regions (non-coding sequences), namely intronic DNA of the ribosomal RNA (both small subunit (18S) and large subunit (28S)). The sequencing results showed *P. brassicae* populations from the *C. occulta* weed carried an authentic intronic pattern and structure. The *Cardamine* group rDNA introns have lost multiple introns in the small and large subunit of rDNA. Moreover, the retained introns despite a largely mutual share of conserved parts with the cosmopolitan strains contained numerous novel structures. These structural differences comprise of high level of polymorphisms such as transversion point mutations occurring at sites involving with the intronic splicing sites or insertions/deletions at the binding sites. Two geographical *P. brassicae* populations from *C. occulta* carried a lengthy intron-encoded ORFs and putative mobile elements established in the large subunit. A few aberrant *P. brassicae* populations from the *Brassica* crops also harbored polymorphic introns that shared common mutated motifs with the weed-affecting group. The diversity of ribosomal introns observed from

those investigated populations also demonstrated the genetic distinction of the *P. brassicae* populations from the *C. occulta*.

This study strongly suggests that *P. brassicae* populations causeing clubroot disease in *C. occulta* are genetically distinct from those causing the disease in vegetables. It was also inferred that coevolution between *P. brassicae* and *C. occulta* has been associated with the dynamics of pathogenicity and genetic diversity of *P. brassicae*.

日本の収穫後水田においてアブラナ科雑草*Cardamine occulta*に根こぶ病を引きおこす*Plasmodiophora brassicae*の性状

摘要

アブラナ科作物に大きな被害をもたらしている根こぶ病は、絶対寄生性の原生動物*Plasmodiphora brassicae*によって引きおこされる土壌伝染病である。根こぶ病の研究は、アブラナ科野菜根こぶ病を対象としたものが主流であるが、日本ではイネ収穫後の水田に自生するアブラナ科雑草のタネツケバナ*Cardamine occulta*にも根こぶ病が発生することが知られ、その病原体（*P. brassicae Cardamine* グループ）も報告されている。タネツケバナは、越冬性のアブラナ科の多年草で、そのゲノムとライフサイクルは農業活動などの人間活動と共進化したものである。タネツケバナ は、8倍体のゲノムを持ち、遺伝的変異に富む。このような特徴は、この雑草が新しい生態的ニッチ、特に水田や植物苗の生産ハウスなどの人為的な生息環境にうまく適応するのに役立っている。同時に、宿主-病原体相互作用を介した雑草の遺伝的変異は、*P. brassicae*個体群の遺伝的多様化を促進させている可能性がある。本研究では、タネツケバナ根こぶ病の病理学的特徴を明らかにするとともに、*P. brassicae Cardamine* groupの分子生物学的性状を解析した。

まず、タネツケバナ根こぶ病の形態学的および病理組織学的特徴について、ブラナ科野菜根こぶ病と比較しながら調査した。タネツケバナ根こぶ病の形態学的特徴として、茎に大きな肥大（茎こぶ）が観察されることや、組織内に多くの*P. brassicae*（一次変形体および休眠胞子）が観察されているにもかかわらず肥大しない場合もあることがわかった。また、顕微鏡観察では、維管束や髄組織、さらには葉柄などの葉面構造に一次変形体が検出された。これらの形態学的特徴は、アブラナ科野菜根こぶ病では見られない特徴であった。

次に、*P. brassicae Cardamine* グループの遺伝的同一性の検証を試みた。検証には、*P. brassicae Cardamine* グループおよび海外で離された*P. brassicae*を含む野菜根こぶ病由来の*P. brassicae* のアクチン遺伝子およびrDNAのITS領域の塩基配列を用いた。

アクチン遺伝子の解析の結果、*P. brassicae* Cardamineグループは、野菜由来の*P. brassicae*のアクチン遺伝子と比較して、7つの点変異（3つのサイレントと4つのミスセンス）を蓄積していることがわかった。

ITS領域の解析の結果、*P. brassicae* CardamineグループはITS1およびITS2領域ともに多くの変異を獲得していることが示された。とくに、ITS2領域では、ITS1領域よりも多くの変異（一塩基多型、オリゴヌクレオチド多型、および挿入・欠失）が見られ、これらの変異は野菜由来の*P. brassicae*よりも多いことがわかった。アクチン遺伝子とITS領域の両塩基配列を連結して系統解析を行った結果、*P. brassicae* Cardamineグループは、野菜由来の*P. brassicae*とは別のグループに属した。また、ITS2領域の塩基配列に基づいて構造解析を行った結果、*P. brassicae* CardamineグループのITS2は、塩基環、バルジとループからなる4つのヘリックスの大きさが野菜由来の*P. brassicae*とは大きく異なっていた。

さらに、rDNA小サブユニット（18S）および大サブユニット（28S）遺伝子のイントロンの配列について、*P. brassicae* Cardamineグループと野菜由来の*P. brassicae*とで比較した。その結果、*P. brassicae* Cardamineグループは、野菜由来の*P. brassicae*で知られている基本的なrDNAのイントロンのパターンと構造を有していたが、*Cardamine*グループのrDNAのイントロンは、小サブユニットおよび大サブユニットで複数のイントロンが失われていた。また、*Cardamine*グループは、保存領域の塩基配列は野菜由来の*P. brassicae*とほぼ共通であったが、イントロンに多数の新規構造を含んでいた。これらの構造は、イントロンスプライシング部位に関わる部位で起こるトランスバージョン点変異や結合部位での挿入・欠失などの複雑な多型から構成されていた。興味深いことに、広島県の*Cardamine*グループの2つの*P. brassicae*集団は、大サブユニットのイントロン内にORFおよびトランスポゾン配列を持っていた。このように、rDNA小サブ

ユニットおよび大サブユニット遺伝子のイントロンの塩基配列の差異からも、*P. brassicae* Cardamineグループと野菜由来の*P. brassicae*を区別できることが示された。

本研究により、タネツケバナに根こぶ病を引きおこす*Plasmodiophora brassicae*は、野菜に根こぶ病を引きおこす*P. brassicae*とは遺伝的に異なることが強く示唆された。また、*P. brassicae*とタネツケバナの間の共進化が*P. brassicae*の病原性や遺伝的多様性に関連してきた可能性が推察された。

LIST OF PUBLICATIONS RELATED TO THE THESIS

1. Lam ATP, Sasaki K, Yanagi Y, Tanaka S, Ito S (2022) Genotypic characterization of *Plasmodiophora brassicae* in the paddy-field weed *Cardamine occulta* and symptomology reveal a distinct pathogen population in Japan. Journal of General Plant Pathology (2022) Published online https://doi.org/10.1007/s10327-022-01106-0 (related to Chapter 2 and 3)

2. Lam ATP, Sasaki K, Yanagi Y, Tanaka S, Ito S (2022) Intraspecific variability and distribution difference within the ribosomal introns of the discrete *Plasmodiophora brassicae* group in Japan: a case study for complex dynamics of intron evolution. Agronomy 12(9):2154. https://doi.org/10.3390/agronomy12092154 (related to Chapter 4)

ACKNOWLEDGEMENT

Words are ineffable for my gratitude towards Japan and the Japanese Government who offered me MONBUKAGAKU-SHO (Ministry of Education, Science, Sports and Culture), a wonderful opportunity to have explored, learned and accomplished myself. Indeed, it grants me so many memorable experiences and broadens my eyes and ears to a whole new world.

Having said that without mentioning my Sensei Shinichi Ito, is a huge mistake. For his wise guidance from my early-year to fully-fledged of doctoral research, I should always feel grateful about that. I would also like to extend my appreciation to my co-supervisor Dr. Yanagi Yukiko, and Dr. Shuhei Tanaka, Dr. Kazunori Sasaki as well as other labmates and students from Plant Pathology lab who kindly supported me when in need.

And obviously the kindest assistance I have received unwaveringly from the staff of both schools Yamaguchi and Totteri University.

For my family, of course they are endless sources of my comfort and support. Thank you! mom, dad, granma.

For my friends, particularly for the Japanese friends who helped me in personal life when I have been in crisis. I would express my greatest attitude to Akiko-san, Yumiko-san, Tomomi-san, Akemi-san, Takematsu Sensei….and a lot of other wonderful Japanese who made my days, despite during the pandemic, so beautifully in Japan.

And for those Vietnamese friends who are always by my side to have cheered me up when I was down. Ngoc, Duy, Tran, Phuong thanks for being my good friends.