

Instytut Chemii Bioorganicznej
Polskiej Akademii Nauk
w Poznaniu



mgr Katarzyna Nowis

Podjęcia bioinformatyczne w badaniach kolistych RNA u roślin

Praca doktorska
wykonana pod kierunkiem
prof. dr. hab. Marka Figlerowicza
Promotor pomocniczy:
dr hab. Anna Philips
w Zakładzie Biologii Molekularnej
i Systemowej
Instytutu Chemii Bioorganicznej
PAN w Poznaniu

Poznań, 2022 r.

*Pragnę podziękować mojemu Promotorowi,
Panu prof. dr hab. Markowi Figlerowiczowi
za opiekę, przekazaną wiedzę, szereg cennych rad,
liczne rozmowy i cierpliwość.*

*Serdecznie dziękuję
mojemu Promotorowi Pomocniczemu,
dr hab. Annie Philips
za nieocenione wsparcie w prowadzonych badaniach,
poświęcony czas na liczne dyskusje,
nieustanną gotowość do pomocy,
cierpliwość oraz tworzenie sprzyjającej atmosfery pracy.*

*Dziękuję również
dr hab. Paulinie Jackowiak
za poświęcony czas
i za chęć dzielenia się swoją ekspercką wiedzą biologiczną.*

*Dziękuję także wszystkim koleżankom i kolegom z
Europejskiego Centrum Bioinformatyki i Genomiki oraz Zakładu Biologii
Molekularnej i Systemowej za burze mózgów, okazaną życzliwość i atmosferę.*

*Pracę dedykuję mojej wspaniałej rodzinie i przyjaciołom w podziękowaniu za
cierpliwość, wiarę i wsparcie.*

Przedłożona praca doktorska została sfinansowana z projektu badawczego
Narodowego Centrum Nauki SONATA 8
“Identyfikacja kolistych RNA oraz białek uczestniczących w ich biogenezie
u modelowej rośliny, *Arabidopsis thaliana*”
nr 2014/15/D/NZ2/02305

SPIS TREŚCI

LISTA PUBLIKACJI	7
wchodzących w skład rozprawy doktorskiej	7
niewchodzących w skład rozprawy doktorskiej	8
STRESZCZENIE	9
ABSTRACT	11
KRÓTKIE OMÓWIENIE PRACY DOKTORSKIEJ	13
Wprowadzenie	13
Cel Pracy	18
Wyniki badań	19
Opracowanie protokołów bioinformatycznych umożliwiających jakościową oraz ilościową analizę circRNA na podstawie danych uzyskanych z sekwencjonowania transkryptomów roślinnych	19
Zastosowanie stworzonych protokołów bioinformatycznych do analizy circRNA powstających w modelowej roślinie <i>A. thaliana</i>	23
Zastosowanie stworzonych protokołów bioinformatycznych do identyfikacji białek zaangażowanych w biogenezę circRNA u <i>A. thaliana</i>	24
Stworzenie bazy danych circRNA u <i>A. thaliana</i>	26
Podsumowanie i perspektywy	27
Bibliografia	29
PUBLIKACJE WCHODZĄCE W SKŁAD ROZPRAWY DOKTORSKIEJ	34
OŚWIADCZENIA DOTYCZĄCE PRAC WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ	88

LISTA PUBLIKACJI

wchodzących w skład rozprawy doktorskiej

1. Philips A, **Nowis K**, Stelmaszczuk M, Jackowiak P, Podkowiński J, Handschuh L, Figlerowicz M;
Expression Landscape of circRNAs in Arabidopsis thaliana Seedlings and Adult Tissues
Frontiers in Plant Science (2020), <https://doi.org/10.3389/fpls.2020.576581>
Pięcioletni IF: 7,3, MEiN 100
2. Philips A, **Nowis K**, Stelmaszczuk M, Podkowiński J, Handschuh L, Jackowiak P, Figlerowicz M;
Arabidopsis thaliana cbp80, c2h2, and flk Knockout Mutants Accumulate Increased Amounts of Circular RNAs
Cells (2020), <https://doi.org/10.3390/cells9091937>
Pięcioletni IF 7,7, MEiN 140
3. **Nowis K**, Jackowiak P, Figlerowicz M, Philips A;
At-C-RNA database, a one-stop source for information on circRNAs in Arabidopsis thaliana in a unified format
Database: The Journal of Biological Databases and Curation (2021), <https://doi.org/10.1093/database/baab074>
Pięcioletni IF 4,8, MEiN 100

niewchodzących w skład rozprawy doktorskiej

1. Czubak K, Taylor K, Piasecka A, Sobczak K, **Kozłowska K**, Philips A, Sedehizadeh S, Brook JD, Wojciechowska M, Kozłowski P;

Global Increase in Circular RNA Levels in Myotonic Dystrophy

Frontiers in Genetics (2019), <https://doi.org/10.3389/fgene.2019.00649>

Pięcioletni IF 4,9, MEiN 100

2. Makowska N, Philips A, Dabert M, **Nowis K**, Trzebny A, Koczura R, Mokracka J;

Metagenomic analysis of β -lactamase and carbapenemase genes in the wastewater resistome

Water Research (2019), <https://doi.org/10.1016/j.watres.2019.115277>

Pięcioletni IF 13,8, MEiN 140

3. Philips A, Stolarek I, Handschuh L, **Nowis K**, Juras A, Trzciński D, Nowaczewska W, Wrzesińska A, Potempa J, Potempa J, Figlerowicz M;

Analysis of oral microbiome from fossil human remains revealed the significant differences in virulence factors of modern and ancient *Tannerella forsythia*

BMC Genomics (2020), <https://doi.org/10.1186/s12864-020-06810-9>

Pięcioletni IF 4,9, MEiN 140

4. Makowska N, Bresa K, Koczura R, Philips A, **Nowis K**, Mokracka J;

Urban wastewater as a conduit for pathogenic Gram-positive bacteria and genes encoding resistance to β -lactams and glycopeptides

Science of the Total Environment (2021),
<https://doi.org/10.1016/j.scitotenv.2020.144176>

Pięcioletni IF 10,2, MEiN 200

STRESZCZENIE

Koliste cząsteczki RNA (circRNA, ang. *circular RNA*) zostały odkryte już w latach 70. ubiegłego wieku. W przypadku roślin był to wiroidowy RNA (Sanger i in., 1976) a w przypadku człowieka transkrypt występujący w cytoplazmie linii komórkowej HeLa (Hsu i in., 1979). Jednakże, przez długi czas circRNA nie cieszyły się większym zainteresowaniem badaczy, gdyż powszechnie sądzono, że ich powstanie jest wynikiem błędów aparatu splicingowego. Sytuacja ta uległa zmianie dopiero na początku XXI wieku wraz z powstaniem technik wysokoprzepustowego sekwencjonowania nowej generacji (NGS, ang. *next-generation sequencing*), które dały pełniejszy wgląd w strukturę transkryptomów. Stwierdzono między innymi, że circRNA są zachowawcze ewolucyjnie i występują powszechnie w organizmach żywych. Obserwacje te pozwalają przypuszczać, że koliste transkrypty pełnią istotne funkcje biologiczne. W rezultacie liczba publikacji dotyczących circRNA zaczęła gwałtownie wzrastać, wciąż jednak niewiele wiadomo na temat biogenezy i funkcji tych cząsteczek szczególnie w odniesieniu do roślin.

Przeprowadzony przegląd literaturowy ujawnił szereg problemów z jakimi zmagają się badacze analizując circRNA u roślin. Pierwszym z nich okazał się brak standardowych protokołów laboratoryjnych i bioinformatycznych dedykowanych analizie circRNA. Powodowało to, że autorzy większości prac jedynie identyfikowali circRNA, nie przeprowadzając porównawczych analiz ich akumulacji w organach czy ekotypach. Kolejnym problemem był brak jednolitego sposobu deponowania informacji na temat roślinnych circRNA w publicznie dostępnych bazach danych. Nie istniała zatem możliwość bezpośredniego porównywania już zidentyfikowanych circRNA oraz poziomów ich akumulacji w różnych organach czy liniach. Jak dotąd dostępne były trzy bazy danych zawierające informacje na temat circRNA u *Arabidopsis thaliana*. Repozytoria te udostępniają informacje na temat circRNA, całkowicie ignorując zagadnienia dotyczące wiarygodności przeprowadzonych analiz jakościowych i ilościowych.

W celu rozwiązania tych problemów opracowałam protokoły bioinformatyczne umożliwiające jakościową i ilościową analizę circRNA występujących w różnych tkankach czy organach roślin. Stworzone protokoły zastosowałam do scharakteryzowania circRNA w kwiatach, liściach, korzeniach i

siewkach *A. thaliana* typu dzikiego (ekotyp Columbia, w skrócie Col-0) oraz do zidentyfikowania białek wpływających na biogenezę circRNA. Uzyskane wyniki wskazują, że u *A. thaliana* większość circRNA powstaje w procesach stochastycznych regulowanych niezależnie od transkrypcji liniowych odpowiedników. Ponieważ circRNA są najprawdopodobniej produktami składania transkryptów, stąd, aby zidentyfikować białka uczestniczące w ich biogenezie scharakteryzowaliśmy circRNA powstające w mutantach *A. thaliana* pozbawionych pojedynczych genów kodujących białka splicingowe. Analizy circRNA w typie dzikim oraz 18 mutantach typu knock-out ujawniły, że wyłączenie trzech genów związanych ze splicingiem wywiera istotny wpływ na produkcję i akumulację circRNA (mutanty: *cbp80*, *flk*, *c2h2*). Brak białek kodowanych przez te geny zaburzał proces składania transkryptów i przyczyniał się do wzrostu produkcji circRNA. Co więcej, w przypadku dwóch wariantów *cbp80*, *c2h2* stwierdziliśmy obecność dużej liczby unikatowych circRNA nie występujących w typie dzikim *A. thaliana*. Dodatkowo, opracowane przeze mnie protokoły dedykowane badaniom circRNA wykorzystałam do analizy publicznie dostępnych danych RNA-seq dla *A. thaliana*. Wszystkie uzyskane wyniki zdeponowałam w stworzonej specjalnie w tym celu publicznej bazie danych (<http://plantcircrna.ibch.poznan.pl/>). Tym samym powstała możliwość jakościowej i ilościowej analizy porównawczej wszystkich circRNA zidentyfikowanych u *A. thaliana*.

ABSTRACT

Circular RNAs (circRNA) were discovered in the 70s of the last century. In the case of plants, it was viroid RNA (Sanger et al. 1976) and in the case of human, the transcript found in the cytoplasm of the HeLa cell line (Hsu et al. 1979). However, little attention was paid to circRNAs for a long time, because it was widely believed that they are products of aberrant splicing. Only the development of next-generation sequencing techniques in the 21st century made it possible to carry out more comprehensive studies of transcriptomes, including circRNA. These studies have revealed that circRNAs are evolutionarily conserved and wild spread in living organisms. These observations suggested that circRNAs have important biological functions. As a result, the number of publications on circRNA began to increase rapidly, but still little is known about the biogenesis and function of these molecules, particularly in plants.

The analysis of articles that have been already published on circRNA in plants revealed a number of problems faced by researchers. The first was the lack of well-validated laboratory and bioinformatics protocols dedicated to circRNA analysis. Most authors identified circRNAs, but did not perform their quantitative comparative analyses in organs or ecotypes. Another problem was the lack of a uniform method for depositing information on plant circRNAs in publicly available databases. As a consequence, it was not possible to directly compare the already collected data on circRNAs occurrence and their accumulation in different plant organs or lines. So far, three databases containing information on *Arabidopsis thaliana* circRNA have been established. These repositories provide information on circRNA identified by using different protocols and completely ignore the issues related to the reliability of the qualitative and quantitative analyzes performed.

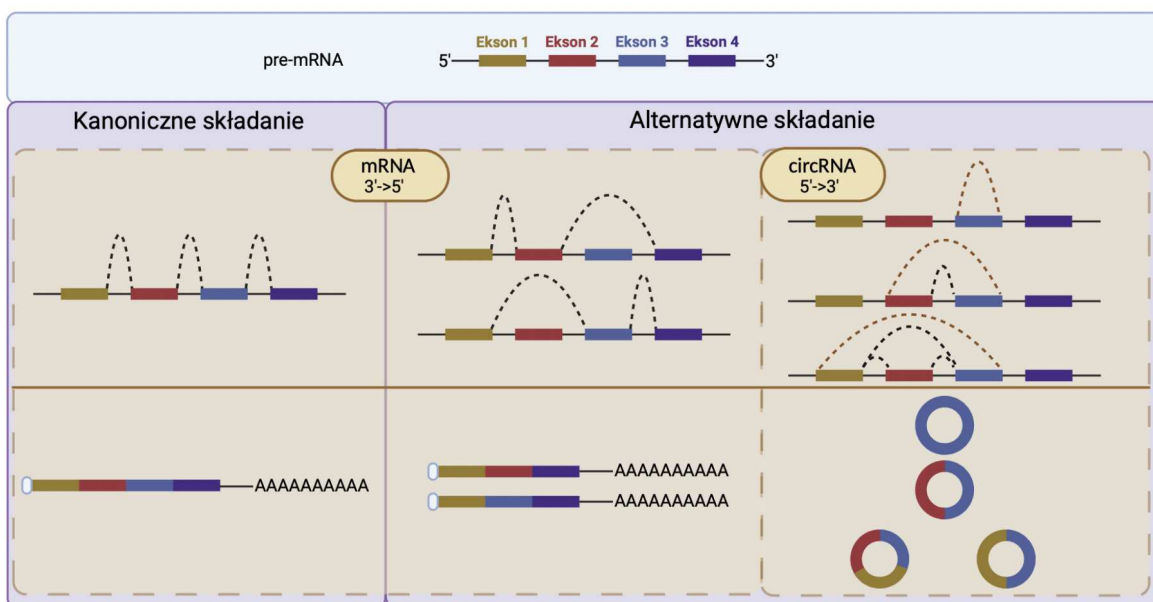
In order to solve these problems, I have developed bioinformatic protocols that enable qualitative and quantitative analyses of circRNAs in various plant tissues or organs. Then I used these protocols to characterize circRNAs in the wild-type *A. thaliana* flowers, leaves, roots, and seedlings (ecotype Columbia, abbreviated as Col-0) and to identify splicing proteins that are involved in the biogenesis of circRNAs. The obtained results indicate that in *A. thaliana* most of the circRNAs are products of stochastic processes that are regulated

independently of the transcription of the linear counterparts. In the next step, we attempted to identify proteins involved in the biogenesis of circRNA. Considering that they are splicing products, we have analyzed circRNA accumulation in *A. thaliana* mutants lacking individual splicing protein genes. Analysis of wild-type circRNA and 18 knock-out mutants revealed that silencing of three splicing-related genes had a significant impact on the production and accumulation of circRNAs (mutants: cbp80, flk, c2h2). It suggests that the lack of proteins encoded by these genes disrupts the transcript processing and contributes to an increase in circRNA production. Moreover, for two variants i.e. cbp80, c2h2, a large number of unique circRNAs not found in wild-type *A. thaliana* was identified. Additionally, I used previously developed circRNA dedicated protocols to analyze publicly available RNA-seq data for *A. thaliana*. All the obtained results were deposited in a specially created public database (<http://plantcircrna.ibch.poznan.pl/>). Thus, the possibility of qualitative and quantitative comparative analyses of all circRNAs identified in *A. thaliana* was created.

KRÓTKIE OMÓWIENIE PRACY DOKTORSKIEJ

Wprowadzenie

RNA jest nie tylko produktem pośrednim w procesie biosyntezy białek. Obok kodujących RNA (mRNA, ang. *messenger RNA*) powstaje także szerokie spektrum różnego typu niekodujących RNA (ncRNA, ang. *noncoding RNA*). W ostatniej dekadzie jako jedną z klas ncRNA wyróżniono koliste RNA (circRNA, ang. *circular RNA*). Są one produktem alternatywnego składania transkryptów, w którym koniec 5' eksonu łączy się kowalencyjnie z końcem 3' tego samego lub jednego z następujących po nim eksonów tworząc zamkniętą pętlę (Rysunek 1). Początkowo uważano, że circRNA powstają w wyniku błędów popełnianych przez maszynę składającą transkrypty. Z tego powodu spodziewano się, że circRNA występują w komórkach w bardzo niskich stężeniach i nie pełnią żadnych funkcji. Dekadę temu okazało się jednak, że circRNA występują powtarzalnie, powszechnie i są zachowawcze ewolucyjnie (Jeck i in., 2012). Kolistą formą tych transkryptów powoduje, iż są one bardziej stabilne niż liniowy RNA. Brak końców 3' i 5' uniemożliwia ich degradację przez liczne komórkowe egzonukleazy (Suzuki i in., 2006).



Rys 1. Podczas składania transkryptów z jednego genu powstają różne formy mRNA oraz circRNA.

W licznych publikacjach wykazano, że circRNA te są typowym elementem komórek eukariotycznych (Jeck i in., 2012; Fan i in., 2015; Shen i in., 2016; Lu i in., 2015; Ye i in., 2015; Huo i in., 2018).

Powyższe obserwacje pozwalają przypuszczać, że circRNA posiadają istotny potencjał funkcjonalny w wielu procesach komórkowych, zwłaszcza w regulacji ekspresji genów (Ashwal-Fluss i in., 2014). Niestety, jak dotąd tylko niewielkiej liczbie circRNA zidentyfikowanych u ludzi i zwierząt udało się przypisać funkcję. Pokazano, że circRNA mogą specyficznie wiązać białka (Ashwal-Fluss i in., 2014; Du i in., 2017; Memczak i in., 2013) lub mikro RNA (miRNA, ang. *micro RNA*) (Hansen i in., 2013; Memczak i in., 2013; Peng i in., 2017). Na przykład circRNA CDR1as/ciRS-7 jest tak zwaną gąbką wiążącą miRNA-7, gdyż ma wiele miejsc do niego komplementarnych. Tym samym circRNA CDR1as/ciRS-7 może wpływać na funkcje miRNA-7, który zaangażowany jest m.in. w regulację procesów nowotworowych (Uhr i in., 2018), a także bierze udział w regulacji ekspresji genu α -synukleiny, białka którego akumulacja w neuronach jest powiązana z szeregiem chorób neurodegeneracyjnych m.in. z chorobą Parkinsona (Lasda i in., 2014; Hansen i in., 2013). Co więcej, wykazano, że circRNA biorą także udział w patogenezie chorób nowotworowych takich jak: nowotwór piersi, prostaty, szyjki macicy, pęcherza, żołądka, wątroby i wielu innych (Liu i in., 2019; Patop i in., 2018).

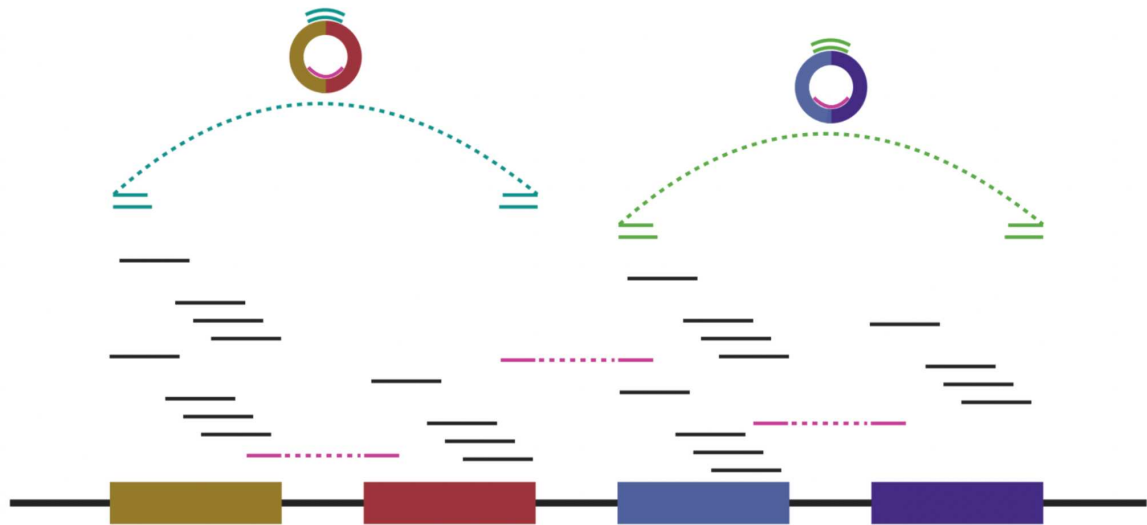
Do tej pory zaproponowano kilka mechanizmów powstawania circRNA (Jeck i in., 2012; Salzman i in., 2012; Salzman i in., 2013). Większość z nich postuluje, że są one produktem alternatywnego splicingu. Podjęto także próby identyfikacji czynników promujących powstawanie circRNA. Zaproponowano między innymi, że na powstawanie circRNA wpływa w pewnym stopniu długość i sekwencja intronów (Jeck i in., 2012; Liang i in., 2014), a także obecność białek wiążących się do RNA (RBP, ang. *RNA binding proteins*) (Ashwal-Fluss i in., 2014; Conn i in., 2015). Wiadomo również, że circRNA generowane są w sposób specyficzny dla komórek/tkanek (Ye i in., 2017). Wymienione powyżej fakty pozwalają postawić hipotezę, że istnieją nieznane czynniki, które wspierają/hamują proces biogenezy circRNA. Z uwagi na fakt, że circRNA i mRNA są wytwarzane z tych samych eksonów i przez tę samą maszynę (spliceosom), jest wysoce prawdopodobne, że proces powstawania circRNA

jest elementem większego systemu regulującego ekspresję genów (Ashwal-Fluss i in., 2014).

W większości opublikowanych jak dotąd prac poświęconych circRNA u *A. thaliana* sugeruje się, że podobnie jak i u zwierząt, cząsteczki te akumulowane są w sposób specyficzny dla tkanki i stadium rozwojowego (Sun i in., 2016; Chen i in., 2017; Liu i in., 2017). Pokazano także, że pojawienie się niektórych circRNA związane jest ze specyficzną odpowiedzią na stres cieplny (Pan i in., 2017). Do tej pory tylko jednemu circRNA występującemu u roślin jednoznacznie przypisano funkcję. Jest to circRNA pochodzący z eksonu 6 genu *SEPALLATA3* (*SEP3*) regulujący rozwój kwiatów u *A. thaliana* (Conn i in., 2017).

W przypadku roślin circRNA zostały najlepiej scharakteryzowane u *A. thaliana*. W tym celu powszechnie wykorzystywano uzyskane wcześniej dane pochodzące z sekwencjonowania (NGS) transkryptomów, szczególnie mRNA. Protokoły stosowane do pozyskania tego typu danych zazwyczaj nie były zoptymalizowane pod kątem badań circRNA, których w komórce jest znacznie mniej niż mRNA. Brak podejść dedykowanych analizom circRNA generowało i nadal generuje szereg problemów. Po pierwsze, do identyfikacji circRNA stosowane były bardzo różnorodne protokoły laboratoryjne i programy bioinformatyczne co prowadziło do identyfikacji zbiorów circRNA istotnie różniących się liczebnością - od kilku do kilkudziesięciu tysięcy, w zależności od stosowanego podejścia (Tabela 1). Po drugie, podczas analiz bioinformatycznych rutynowo zakładano, że circRNA został wykryty, jeśli jego obecność jest poparta co najmniej dwoma odczytami odpowiadającymi miejscu splicingu circRNA (ang. *back-splicing site*) (Rysunek 2). Przyjęcie tak łagodnego kryterium bez uwzględnienia innych parametrów wpływających na wynik sekwencjonowania w tym także jego głębokości prowadziło często do uzyskania pozytywnie fałszywych rezultatów identyfikacji circRNA, lub identyfikacji takich które powstały w sposób losowy i są niepowtarzalne. Po trzecie, tylko w kilku przypadkach akumulacja wybranych circRNA zidentyfikowanych na podstawie danych z sekwencjonowania została potwierdzona eksperymentalnie, np. metodą ilościowej łańcuchowej reakcji polimerazy (qPCR, ang. *quantitative polymerase chain reaction*) (Wang i in., 2014; Liu i in., 2017; Cheng i wsp., 2017; Pan i in., 2017). Brak walidacji wyników z RNA-seq za pomocą innych metod często uniemożliwia wykorzystanie danych z sekwencjonowania

do wielokierunkowych, przekrojowych analiz akumulacji circRNA. W związku z tym, jak do tej pory nie udzielono zadowalających odpowiedzi na wiele pytań dotyczących circRNA, w tym poziomu ich akumulacji w roślinach, funkcji i biogenezy.



Rys 2. Identyfikacja circRNA na podstawie odczytów NGS wspierających miejsca splicingu circRNA (5'-3', ang. *back-splicing site*). Kolorami zaznaczone są odczyty: czarnym - zmapowane w całości do eksonów, różowym - zmapowane do kanonicznych miejsc składania (3'-5') będące potwierdzeniem liniowego złożenia eksonów, oraz zielonym - zmapowane w odwrotnej kolejności (5'-3') do poprzedzającego i następującego eksonu będące potwierdzeniem circRNA.

Tabela 1. Liczba circRNA zidentyfikowanych przez różne grupy badawcze (na podstawie danych zdeponowanych w Plantcircbase; Chu i in., 2017).

Organizm	Liczba circRNA	Referencja
<i>Arabidopsis thaliana</i>	6	Wang et al., 2014
	6 012	Ye et al., 2015
	976	Sun et al., 2016
	30 534	Chu et al., 2017
	4 415	Chen et al., 2017
	165	Lu et al., 2017
	2 167	Pan et al., 2017
	168	Liu et al., 2017
	160	Dou et al., 2017
	459	Zhang et al., 2019
	12 663	Philips et al., 2020
	333	Zhang et al., 2020

Cel Pracy

Celem niniejszej rozprawy było opracowanie, a następnie optymalizacja podejść bioinformatycznych umożliwiających ilościową i jakościową analizę circRNA u roślin a następnie zastosowanie tych podejść do:

- (i) identyfikacji circRNA występujących w modelowej roślinie dwuliściennej *A. thaliana* (w siewkach i różnych organach dojrzałej rośliny);
- (ii) identyfikacji białkowych komponentów spliceosomu uczestniczących w biogenezie circRNA u *A. thaliana*;
- (iii) opracowania bazy danych circRNA u *A. thaliana*.

Wyniki badań

1. Opracowanie protokołów bioinformatycznych umożliwiających jakościową oraz ilościową analizę circRNA na podstawie danych uzyskanych z sekwencjonowania transkryptomów roślinnych

Mimo rosnącej liczby publikacji na temat circRNA u *A. thaliana* stan wiedzy o biogenezie i funkcji molekularnej tych RNA jest niewielki. Jednym z głównych powodów jest brak zwalidowanych i wystandaryzowanych protokołów jakościowej i ilościowej analizy circRNA.

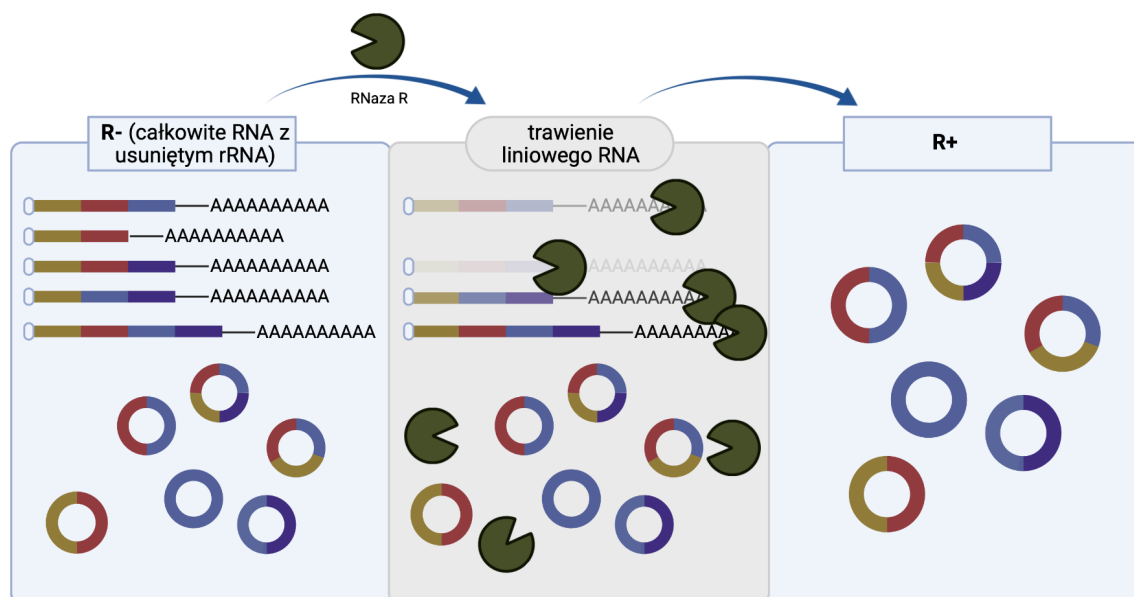
W przypadku pierwszego etapu analizy circRNA, polegającego na uzyskaniu próbek RNA do sekwencjonowania najczęściej stosowane są dwa podejścia: (i) izolacja całkowitego RNA a następnie usunięcie z niego rRNA oraz (ii) izolacja całkowitego RNA, usunięcie rRNA a następnie trawienie RNazą R (Rysunek 3). Usunięcie rRNA jest istotne, ponieważ stanowi on ponad 80% wszystkich transkryptów i powoduje, że wykrycie RNA o stosunkowo niskiej akumulacji (takich jak circRNA) staje się niemal niemożliwe. Dodatkowo, w celu usunięcia liniowych transkryptów próbki traktowane są RNazą R. Enzym ten jest egzorybonukleazą trawiącą RNA w kierunku 3' -> 5', circRNA są zatem odporne na jego działanie.

Jak dotąd stworzono szereg algorytmów służących do identyfikacji circRNA na podstawie danych RNA-seq. Algorytmy te charakteryzują się różną specyficznością i czułością, zatem generują różne zbiory potencjalnych circRNA dla tych samych danych wejściowych. Ponadto w analizach tych powszechnie stosowano standardy normalizacji danych RNA-seq opracowane wcześniej dla badań mRNA. Nikt nie zweryfikował tych metod normalizacji w kontekście analiz circRNA u roślin. W związku z powyższym, aby stworzyć wiarygodny protokół do analizy jakościowej i ilościowej circRNA u roślin musiałam opracować optymalny schemat: przygotowania bibliotek NGS, normalizacji danych RNA-seq oraz identyfikacji circRNA. Wyniki tych badań zawarte są w publikacji 1, której jestem współautorką.

Materiałem wykorzystanym do badań był RNA wyizolowany z siewek, korzeni, liści i kwiatów *A. thaliana* typu dzikiego. Następnie próbki zostały przygotowane do sekwencjonowania NGS. Wszystkie eksperymenty wykonano

w czterech powtórzeniach biologicznych. Do identyfikacji circRNA na podstawie danych z sekwencjonowania zastosowałam program CIRI2 (Gao i in., 2017), który wg. testu przeprowadzonego przez (Zhang i in., 2020) jest najlepszy do identyfikacji circRNA *de novo*, a gdy zastosowany u roślin wykazuje wysoką specyficzność i czułość.

W pierwszym etapie przetestowałam dwa wyżej wspomniane protokoły przygotowania próbek do analiz circRNA: (i) z usunięciem rRNA i traktowaniem RNazą R (próbki oznaczone R+) oraz (ii) z usunięciem rRNA i nietraktowaniem RNazą R (próbki oznaczone R-) (Rysunek 3). Wyniki przeprowadzonych przeze mnie analiz wykazały, że traktowanie próbek RNazą R znacząco zwiększyło liczbę zidentyfikowanych circRNA, jednak spektrum circRNA o wysokiej akumulacji uległo zmniejszeniu. Niektóre z circRNA, występujące w dużym stężeniu, których obecność została potwierdzona metodą PCR, nie były wykrywane w próbkach traktowanych RNazą R. Najprawdopodobniej cząsteczki te ulegały zliniowaniu podczas inkubacji, przez co stawały się celem działania RNazy R. Biorąc powyższe pod rozwagę zdecydowałam, że w celu uzyskania możliwie reprezentatywnej puli circRNA należy zrezygnować ze wzbogacania próbek w circRNA przy użyciu RNazy R. W dalszych analizach wykorzystywałam zatem wyniki sekwencjonowania uzyskane dla próbek R-. Dodatkową zaletą tego podejścia jest możliwość jednoczesnego przeprowadzenia analiz liniowych i kolistych RNA co byłoby niemożliwe w przypadku próbek R+. Niezależnie jednak od zastosowanej strategii przygotowania próbek powtarzalność z jaką circRNA występowały w poszczególnych powtórzeniach biologicznych była niewielka. CircRNA występujące powtarzalnie we wszystkich 4 powtórzeniach biologicznych stanowiły zaledwie od 1,3% (R-) do 2,7% (R+) całkowitej puli circRNA zidentyfikowanych w liściach oraz od 4,4% (R-) do 5,0% (R+) puli circRNA zidentyfikowanych w korzeniach (Rysunek 2 w Publikacji 1). Z tej przyczyny, uznałam, że w dalszych analizach będę brała pod uwagę jedynie te circRNA, które występują we wszystkich 4 powtórzeniach biologicznych.



Rys 3. Przygotowanie próbek do analiz circRNA.

Kolejnym etapem była jakościowa i ilościowa analiza porównawcza akumulacji circRNA w organach i siewkach rośliny *A. thaliana* typu dzikiego. Podstawowym warunkiem rzetelnego przeprowadzenia tego typu analiz jest normalizacja użytych w niej danych. Z doniesień literaturowych na temat circRNA u roślin wynika, że dotychczas stosowano normalizację uwzględniającą wielkość biblioteki. Dodatkowo zebrane dane świadczą, że uzyskiwane wyniki nie były poddawane walidacji innymi technikami laboratoryjnymi (Ye i in. 2015), lub walidacja ograniczała się do pojedynczych circRNA (Chen i in., 2017, Liu i in., 2017). W związku z tym w pierwszej analizie danych z sekwencjonowania (RNA-seq) zastosowałam normalizację przez wielkość biblioteki, a następnie poprawność tego podejścia została zweryfikowana przy pomocy emulsyjnego PCR (ddPCR, ang. *droplet digital PCR*).

Akumulacja losowo wybranych 10 circRNA (zidentyfikowanych wcześniej na podstawie danych RNA-seq) została zmierzona w 3 powtórzeniach biologicznych przy użyciu ddPCR. Uzyskane wyniki pozwoliły stwierdzić, że pomiary stężenia circRNA metodą ddPCR są wiarygodne jedynie wówczas, gdy analizowana cząsteczka występuje w co najmniej 50 kopiach/ μ g całkowitego RNA. Co ciekawe współczynnik korelacji Pearsona pomiędzy akumulacją circRNA wyznaczoną na podstawie danych RNA-seq znormalizowanych przez wielkość biblioteki a akumulacją określoną metodą ddPCR wyniósł zaledwie 0,67.

Ponieważ metoda ddPCR stanowi obecnie tzw. złoty standard w analizie ilościowej kwasów nukleinowych, uznałam że problem leży najprawdopodobniej po stronie analizy danych RNA-seq.

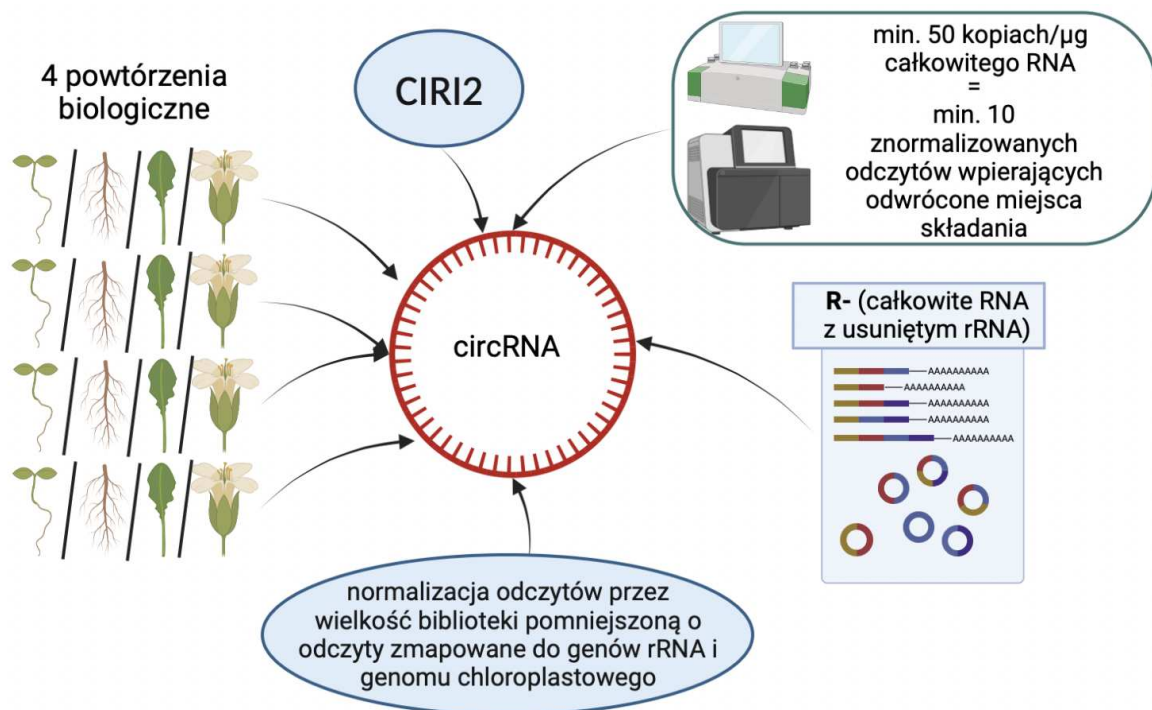
Aby poznać przyczyny umiarkowanej korelacji pomiędzy danymi uzyskanymi metodą RNA-seq (znormalizowanymi przez wielkość biblioteki) oraz danymi otrzymanymi metodą ddPCR przeprowadziłam analizę składu badanych próbek RNA. Zastosowana normalizacja przez wielkość biblioteki wydaje się bowiem sensowna jedynie wówczas, gdy ilość poszczególnych typów transkryptów w próbkach jest zbliżona. Przeprowadzona analiza ujawniła istotne różnice w szczególności w liczbie odczytów zmapowanych do rRNA oraz genomu chloroplastowego (od 0,86% w korzeniach do 6,24% w liściach oraz od 2,58% w korzeniach do 59,51% w liściach odpowiednio dla rRNA i genomu chloroplastowego; Tabela 2 z Suplementu w Publikacji 1). Różnice w liczebności chloroplastowego RNA można tłumaczyć zwiększoną liczbą tych organelli w liściach i siewce w porównaniu do innych kwiata czy korzenia, a rRNA różną wydajnością strategii ich usuwania z organów/siewki. Po wykluczeniu tych odczytów współczynnik korelacji wzrósł do 0,76.

Kolejne metody normalizacji testowane przeze mnie, przez sumę odczytów wspierających circRNA oraz przez współczynnik wyznaczony przez program DESeq2 (Love i in., 2014), dały odpowiednio współczynniki korelacji Pearsona 0,77 i 0,76. Podobną korelację uzyskałam normalizując odczyty przez liczbę odczytów zmapowanych się do genu referencyjnego *ACT2* (0,77), którego poziom ekspresji jest stabilny w całej roślinie. Co więcej, korelacja Pearsona pomiędzy wynikami uzyskanymi dla tych samych organów była wysoka niezależnie od doboru metody normalizacji odczytów (0,88 - kwiat; 0,90 - korzeń; 0,92 - liść). Jedynie dane uzyskane dla siewek wykazały mniejszą korelację (0,76) co można tłumaczyć ogólnie zmniejszoną liczebnością circRNA.

Uzyskane wyniki pozwoliły na wyłonienie czterech metod normalizacji: przez (1) liczbę odczytów wspierających koliste miejsce cięcia, (2) współczynnik wielkości wyznaczony przez DESeq2, (3) wielkość biblioteki bez odczytów z rRNA i chloroplastów oraz (4) liczbę odczytów mapujących się do genu referencyjnego *ACT2*, które dały zbliżone współczynniki korelacji między 0,76 - 0,77. Do dalszych analiz użyłam metody normalizacji odczytów przez wielkość biblioteki pomniejszoną o odczyty zmapowane do genów rRNA i genomu chloroplastowego.

2. Zastosowanie stworzonych protokołów bioinformatycznych do analizy circRNA powstających w modelowej roślinie *A. thaliana*

Zastosowanie opisanej powyżej strategii umożliwiło identyfikację circRNA i wykonanie analiz różnicowej akumulacji circRNA pomiędzy organami/siewką w typie dzikim *A. thaliana* (Rysunek 4, Publikacja 1). Podsumowanie zastosowanych kryteriów zestawiałam na Rysunku 4. W sumie zidentyfikowałam 127 circRNA: 109 w siewce, 87 w korzeniu, 79 w liściu i 99 w kwiecie. Wspólnych we wszystkich organach i siewce było 52 circRNA (Rysunek 3 w Publikacji 1). Porównania akumulacji circRNA parami między organami/siewką wykazały, że circRNA są akumulowane specyficcznie dla organu/siewki. Najwięcej circRNA o istotnie zwiększonej akumulacji w stosunku do innych organów/siewki było w korzeniu w przeciwieństwie do siewki, w której zidentyfikowałam najmniej circRNA o zwiększonej akumulacji w stosunku do badanych organów (Rysunek 4 w Publikacji 1).



Rys 4. Zastosowane kryteria filtrowania w analizie zmierzającej do identyfikacji circRNA o potencjale funkcjonalnym.

Zastosowanie biblioteki całkowitego RNA z usuniętym rRNA i bez użycia RNazy R pozwoliło także sprawdzić, czy produkcja circRNA jest powiązana

z produkcją liniowych odpowiedników. Wyniki wykazały, że nie ma takiej korelacji (Rysunek 5 w Publikacji 1). Porównałam także akumulację circRNA i ich liniowych odpowiedników w organie/siewce na tle produkcji we wszystkich organach i siewce. Liść i siewka wykazały zwiększoną produkcję circRNA w stosunku do liniowych odpowiedników w porównaniu z sumaryczną produkcją we wszystkich organach i siewce (Rysunek 6 w Publikacji 1). Porównanie akumulacji odczytów wspierających circRNA oraz ich liniowych odpowiedników zostało potwierdzone przy użyciu techniki ddPCR (Rysunek 7 w Publikacji 1). Na podstawie poczynionych obserwacji doszliśmy do wniosku, że biogeneza circRNA u *A. thaliana* jest regulowana w sposób specyficzny dla danego organu. Co więcej proces powstawania circRNA jest jednym z elementów złożonego mechanizmu regulacji ekspresji genów.

3. Zastosowanie stworzonych protokołów bioinformatycznych do identyfikacji białek zaangażowanych w biogenezę circRNA u *A. thaliana*

Do tej pory postulowano kilka mechanizmów powstawania circRNA, m.in. związanych z obecnością intronów flankujących, czy z zaangażowaniem białek wiążących RNA. Co więcej, pokazano, że circRNA produkowane są w sposób specyficzny dla komórki. Mimo tego, jak dotąd nie udało się zidentyfikować konkretnych czynników zaangażowanych w cyrkularyzację transkryptów podczas procesu ich składania. Z tego powodu postanowiliśmy zidentyfikować geny, których produkty wpływają na produkcję circRNA.

A. thaliana ma tysiące łatwo dostępnych mutantów, w tym dobrze scharakteryzowane linie z knockoutami pojedynczych genów. Biorąc pod uwagę przypuszczalny mechanizm powstawania circRNA, do dalszych badań obok typu dzikiego *A. thaliana*, wykorzystaliśmy 18 wariantów, z których każdy zawiera mutację typu knockout w genie kodującym białko uczestniczące w splicingu (Wang i Brendel, 2004). Identyfikacja i analiza poziomu akumulacji circRNA przeprowadzona została zgodnie z wcześniej opracowanym i zoptymalizowanym protokołem. Podobnie jak w przypadku roślin typu dzikiego, analiza wariantów pokazała, że tylko niewielka frakcja circRNA jest produkowana powtarzalnie i stąd może posiadać jakiś potencjał funkcjonalny. Ponadto stwierdziliśmy, że dwa z analizowanych wariantów (*cbp80* oraz *c2h2*) produkują istotnie większą liczbę

circRNA w porównaniu ze średnią wyznaczoną dla wszystkich wariantów. Oba te warianty produkują także największą liczbę unikatowych circRNA (odpowiednio: 118 i 61; Rysunek 1 w Publikacji 2). Co więcej, trzy z mutantów (*cbp80*, *c2h2* oraz *flk*) wykazują istotnie zwiększoną akumulację circRNA w stosunku do typu dzikiego (Rysunek 2 i 3 w Publikacji 2). Sugeruje to, że brak białek *cbp80*, *c2h2* oraz *flk* zaburza proces składania transkryptów, co może powodować zwiększoną częstość powstawania circRNA. Biorąc powyższe pod uwagę postanowiłam bliżej przyjrzeć się tym 3 mutantom.

W pierwszym z nich wykluczono ekspresję genu *cbp80*, którego produkt wraz z CBP20 jest istotnym elementem kompleksu wiążącego się z czapeczką (CBC, ang. *cap binding complex*). Jak dotąd wiadomo, że białko CBP80 jest zaangażowane w transdukcję sygnału kwasu abscysynowego (ABA, ang. *abscisic acid*) (Hugouvieux i in., 2002) i szlak kwitnienia (Kuhn i in., 2007; Kuhn i in., 2008). Dodatkowo, mutant pozbawiony CBP80 wykazuje indukowane przez ABA podwyższenie cytozolowego poziomu wapnia w komórkach ochronnych, co prowadzi do zwiększonego zamykania aparatów szparkowych i zapewnienia zwiększoną tolerancję na suszę (Hugouvieux i in., 2002). Mutanty *cbp80* wykazują również fenotyp wczesnego kwitnienia, wynikający z wadliwego składania intronów FLOWERING LOCUS C (FLC), co zmniejsza poziom prawidłowo składanych transkryptów FLC (Kuhn i in., 2007). Sam CBC jest również zaangażowany w biogenezę miRNA, a mutanty w których wyciszono ekspresję CBC, albo jego komponenty akumulują częściowo złożone transkrypty (Laubinger i in., 2008) oraz pri-miRNA (Kim i in., 2008). Co ciekawe, brak CBC, a w szczególności CBP80 wpływa także na alternatywny splicing, szczególnie na wybór miejsca splicingu na końcu 5' pierwszego intronu (Raczyńska i in., 2009). Moje wyniki są spójne z powyższymi odkryciami. Brak CBP80 preferencyjnie wpływa na proces składania w regionie mRNA bliższym końca 5'. Wykazałam, że budowa circRNA powstających w mutancie *cbp80* jest inna niż w pozostałych mutantach oraz typie dzikim. Brak CBP80 sprzyjał cyrkularyzacji transkryptów począwszy od 1. eksonu. W odróżnieniu od tego mutantu, *c2h2*, *flk* i typ dziki produkowały circRNA ulegające cyrkularyzacji głównie od 2. lub kolejnych eksonów (Patrz Rysunek 4 w Publikacji 2).

W drugim mutancie, *c2h2*, zahamowana została ekspresja genu kodującego białko współtworzące kompleks U4/U6.U5 tri-snRNP, który odgrywa

kluczową rolę w tworzeniu aktywnego katalitycznie spliceosomu. W trzecim mutancie *flk*, nieobecne było białko uczestniczące w rozpoznawaniu nie w pełni dojrzałych pre-mRNA, które nie są gotowe i z tego powodu nie powinny być eksportowane do cytoplazmy. W przypadku żadnego z wariantów (w tym w typie dzikim) nie występowała korelacja pomiędzy akumulacją transkryptów liniowych i circRNA. Powyższa obserwacja pozwala postawić wniosek, że produkcja liniowych i kolistych RNA jest regulowana niezależnie.

Analiza funkcjonalna genów (GO, ang. *Gene Ontology*) z których powstawały unikatowe circRNA w wariantach *cbp80* i *c2h2* wykazała istotne statystycznie zwiększenie udziału genów zaangażowanych w odpowiednio odpowiedź na stres i procesy komórkowe.

4. Stworzenie bazy danych circRNA u *A. thaliana*

Przegląd artykułów dotyczących identyfikacji i różnicowej akumulacji circRNA w roślinach wykazał, że opublikowane wyniki są w dużej mierze niespójne z powodu braku wystandaryzowanych metod identyfikacji oraz ilościowej analizy circRNA. Baza danych PlantcircBase zawierająca circRNA zidentyfikowane w roślinach została utworzona na podstawie publicznie dostępnych zasobów wytworzonych przez różne grupy badawcze różnymi metodami eksperymentalnymi i bioinformatycznymi (Chu i in., 2017). Niestety, ze względu na znaczne różnice w zastosowanych procedurach danych tych nie można bezpośrednio porównywać. Na przykład, liczba circRNA znalezionych w *A. thaliana* waha się od 6 (Wang i in., 2014) do 30 534 (Chu i in., 2017), w zależności od podejścia eksperymentalnego (Tabela 1). Dzięki opracowanemu protokołowi (Publikacja 1) byłam w stanie ponownie przeanalizować tzw. surowe ogólnodostępne dane z sekwencjonowania całkowitego RNA (pozbawionego rRNA) u *A. thaliana*. Efektem tych działań jest baza prezentująca zbiór circRNA zidentyfikowanych w ten sam sposób na podstawie naszych oraz innych ogólnie dostępnych danych RNA-seq. W bazie tej podano także znormalizowany poziom akumulacji poszczególnych circRNA, dzięki czemu można go porównywać pomiędzy organami, liniami i różnymi eksperymentami. Baza danych At-C-RNA została opisana w publikacji 3 i jest aktualnie największym

źródłem informacji na temat circRNA u *A. thaliana*. Baza danych jest dostępna pod adresem: <http://plantcircrna.ibch.poznan.pl/>.

Podsumowanie i perspektywy

Publikacje wchodzące w skład niniejszej rozprawy doktorskiej prezentują wyniki pierwszej systematycznej analizy circRNA u *A. thaliana*. Ukazują one sposób w jaki rozwiązany został problem analizy jakościowej i ilościowej circRNA u *A. thaliana* i roślin w ogóle. Dzięki opracowanym procedurom zidentyfikowano circRNA powstające w 3 organach oraz siewce *A. thaliana*. Następnie przeprowadzono ilościową analizę porównawczą akumulacji circRNA w poszczególnych organach/siewce (Publikacja 1). Ponadto, analiza 18 wariantów *A. thaliana*, z których każdy pozbawiony był jednego z białek uczestniczących w procesie składania pre-mRNA (mutanty typu knockout) umożliwiła identyfikację 3 białek, których brak istotnie wpływał na wzrost liczby oraz poziomu akumulacji circRNA (Publikacja 2). Dodatkowo, uzyskane wyniki oraz wyniki wcześniejszych badań poddane zostały standaryzacji i udostępnione w publicznie dostępnej bazie danych (Publikacja 3).

Rosnąca liczba badań dotyczących circRNA u roślin pozwoliła wysnuć wnioski, że circRNA są produkowane specyficznym dla danego typu komórek, tkanek, stadium rozwojowego czy w odpowiedzi na stres np. suszy (Zhang i in., 2020). Obserwacje te sugerują, że circRNA mogą być post transkrypcyjnymi czynnikami regulującymi ekspresję genów. Co więcej, wykazano także u zwierząt, że poszczególne circRNA mogą mieć potencjał do kodowania białek (Legnini i in., 2017; Yang i in., 2017). Pomimo tego, biogeneza i potencjał funkcjonalny circRNA u roślin pozostaje nieznany. Prace wchodzące w skład niniejszej rozprawy stanowią solidny fundament do dalszych badań circRNA u roślin. W najbliższym czasie powinny one skoncentrować się na poznaniu zarówno molekularnych mechanizmów leżących u podłoża biogenezy circRNA jak i funkcji jakie pełnią te cząsteczki. Ważnym zadaniem jest też poznanie pełnych sekwencji circRNA, których odczytanie umożliwiają techniki sekwencjonowania trzeciej generacji (m. in. nanopore). Niedawno Xin i inni (2021) opracowali metodykę laboratoryjną i bioinformatyczną, którą z powodzeniem wykorzystali

do odczytania pełnych sekwencji transkryptów circRNA u człowieka przy użyciu sekwencjonowania nanopore. Dostosowanie tego protokołu do analizy circRNA u roślin, a w dalszej perspektywie zmierzenie ich akumulacji na podstawie pełnych sekwencji (a nie tylko na podstawie odczytów wspierających miejsca splicingu circRNA) mogłyby istotnie przyspieszyć badania nad biogenezą i funkcją circRNA.

Ciekawym aspektem byłoby także zbadanie wpływu kolistych RNA na fenotyp poszczególnych mutantów. Mutant *cbp80* wydaje się być ciekawym kandydatem do tego typu badań. Jego fenotyp został wcześniej dobrze scharakteryzowany. Dodatkowo z przeprowadzonych badań wynika, że brak białka CBP80 prowadzi do istotnych zmian w liczebności i poziomie akumulacji circRNA w porównaniu z typem dzikim. Poszerzenie badań nad mutantem *cbp80* w kontekście circRNA i miRNA mogłoby ujawnić możliwe mechanizmy regulacji dojrzewania miRNA przez circRNA. Kuszącą jest też hipoteza, że być może miRNA w jakiś sposób inhibują powstawanie circRNA, a ich brak wpływa na zwiększenie akumulacji circRNA.

Bibliografia

- Ashwal-Fluss, Reut, et al. "CircRNA Biogenesis Competes with Pre-mRNA Splicing." *Molecular Cell*, vol. 56, no. 1, 2014, pp. 55–66., doi:10.1016/j.molcel.2014.08.019.
- Chen, Gang, et al. "Genome-Wide Identification of Circular RNAs in Arabidopsis Thaliana." *Frontiers in Plant Science*, vol. 8, 2017, doi:10.3389/fpls.2017.01678.
- Cheng, Jinping, et al. "A Lariat-Derived Circular RNA Is Required for Plant Development in Arabidopsis." *Science China Life Sciences*, vol. 61, no. 2, 2017, pp. 204–213., doi:10.1007/s11427-017-9182-3.
- Chu, Qinjie, et al. "PlantcircBase: A Database for Plant Circular RNAs." *Molecular Plant*, vol. 10, no. 8, 2017, pp. 1126–1128., doi:10.1016/j.molp.2017.03.003.
- Conn, Simon J., et al. "The RNA Binding Protein Quaking Regulates Formation of CircRNAs." *Cell*, vol. 160, no. 6, 2015, pp. 1125–1134., doi:10.1016/j.cell.2015.02.014.
- Conn, Vanessa M., et al. "A CircRNA from SEPALLATA3 Regulates Splicing of Its Cognate mRNA through R-Loop Formation." *Nature Plants*, vol. 3, no. 5, 2017, doi:10.1038/nplants.2017.53.
- Du, William W, et al. "Identifying and Characterizing CircRNA-Protein Interaction." *Theranostics*, vol. 7, no. 17, 2017, pp. 4183–4191., doi:10.7150/thno.21299.
- Fan, Xiaoying, et al. "Single-Cell RNA-Seq Transcriptome Analysis of Linear and Circular RNAs in Mouse Preimplantation Embryos." *Genome Biology*, vol. 16, no. 1, 2015, doi:10.1186/s13059-015-0706-1.
- Gao, Yuan, et al. "Circular RNA Identification Based on Multiple Seed Matching." *Briefings in Bioinformatics*, vol. 19, no. 5, 2017, pp. 803–810., doi:10.1093/bib/bbx014.
- Gregory, Brian D., et al. "A Link between RNA Metabolism and Silencing Affecting Arabidopsis Development." *Developmental Cell*, vol. 14, no. 6, 2008, pp. 854–866., doi:10.1016/j.devcel.2008.04.005.
- Hansen, Thomas B., et al. "Natural RNA Circles Function as Efficient MicroRNA Sponges." *Nature*, vol. 495, no. 7441, 2013, pp. 384–388.,

doi:10.1038/nature11993.

Hsu, Ming-Ta, and Miguel Coca-Prados. "Electron Microscopic Evidence for the Circular Form of RNA in the Cytoplasm of Eukaryotic Cells." *Nature*, vol. 280, no. 5720, 1979, pp. 339–340., doi:10.1038/280339a0.

Hugouvieux Véronique, et al. "Localization, Ion Channel Regulation, and Genetic Interactions during Abscisic Acid Signaling of the Nuclear mRNA Cap-Binding Protein, ABH1." *Plant Physiology*, vol. 130, no. 3, 2002, pp. 1276–1287., doi:10.1104/pp.009480.

Huo, Liang, et al. "Genome-Wide Identification of CircRNAs in Pathogenic Basidiomycetous Yeast *Cryptococcus Neoformans* Suggests Conserved CircRNA Host Genes over Kingdoms." *Genes*, vol. 9, no. 3, 2018, p. 118., doi:10.3390/genes9030118.

Jeck, William R., et al. "Circular RNAs Are Abundant, Conserved, and Associated with ALU Repeats." *Rna*, vol. 19, no. 2, 2012, pp. 141–157., doi:10.1261/rna.035667.112.

Kim, S., et al. "Two Cap-Binding Proteins CBP20 and CBP80 Are Involved in Processing Primary MicroRNAs." *Plant and Cell Physiology*, vol. 49, no. 11, 2008, pp. 1634–1644., doi:10.1093/pcp/pcn146.

Kuhn, J. M., et al. "MRNA Cap Binding Proteins: Effects on Abscisic Acid Signal Transduction, MRNA Processing, and Microarray Analyses." *Current Topics in Microbiology and Immunology Nuclear Pre-MRNA Processing in Plants*, 2008, pp. 139–150., doi:10.1007/978-3-540-76776-3_8.

Kuhn, Josef M., et al. "MRNA Metabolism of Flowering-Time Regulators in Wild-Type Arabidopsis Revealed by a Nuclear Cap Binding Protein Mutant, *abh1*." *The Plant Journal*, vol. 50, no. 6, 2007, pp. 1049–1062., doi:10.1111/j.1365-313x.2007.03110.x.

Lasda, Erika, and Roy Parker. "Circular RNAs: Diversity of Form and Function." *Rna*, vol. 20, no. 12, 2014, pp. 1829–1842., doi:10.1261/rna.047126.114.

Laubinger, S., et al. "Dual Roles of the Nuclear Cap-Binding Complex and SERRATE in Pre-MRNA Splicing and MicroRNA Processing in Arabidopsis Thaliana." *Proceedings of the National Academy of Sciences*, vol. 105, no. 25, 2008, pp. 8795–8800., doi:10.1073/pnas.0802493105.

Laubinger, S., et al. "Dual Roles of the Nuclear Cap-Binding Complex and

SERRATE in Pre-mRNA Splicing and MicroRNA Processing in Arabidopsis Thaliana.” *Proceedings of the National Academy of Sciences*, vol. 105, no. 25, 2008, pp. 8795–8800., doi:10.1073/pnas.0802493105.

Legnini, Ivano, et al. “Circ-ZNF609 Is a Circular RNA That Can Be Translated and Functions in Myogenesis.” *Molecular Cell*, vol. 66, no. 1, 2017, doi:10.1016/j.molcel.2017.02.017.

Liang, Dongming, and Jeremy E. Wilusz. “Short Intronic Repeat Sequences Facilitate Circular RNA Production.” *Genes & Development*, vol. 28, no. 20, 2014, pp. 2233–2247., doi:10.1101/gad.251926.114.

Liu, Jianhong, et al. “Circular RNAs: The Star Molecules in Cancer.” *Molecular Aspects of Medicine*, vol. 70, 2019, pp. 141–152., doi:10.1016/j.mam.2019.10.006.

Liu, Tengfei, et al. “Identifying and Characterizing the Circular RNAs during the Lifespan of Arabidopsis Leaves.” *Frontiers in Plant Science*, vol. 8, 2017, doi:10.3389/fpls.2017.01278.

Love, Michael I, et al. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology*, vol. 15, no. 12, 2014, doi:10.1186/s13059-014-0550-8.

Lu, Tingting, et al. “Transcriptome-Wide Investigation of Circular RNAs in Rice.” *Rna*, vol. 21, no. 12, 2015, pp. 2076–2087., doi:10.1261/rna.052282.115.

Memczak, Sebastian, et al. “Circular RNAs Are a Large Class of Animal RNAs with Regulatory Potency.” *Nature*, vol. 495, no. 7441, 2013, pp. 333–338., doi:10.1038/nature11928.

Pan, Ting, et al. “Heat Stress Alters Genome-Wide Profiles of Circular RNAs in Arabidopsis.” *Plant Molecular Biology*, vol. 96, no. 3, 2017, pp. 217–229., doi:10.1007/s11103-017-0684-7.

Patop, Ines Lucia, and Sebastian Kadener. “CircRNAs in Cancer.” *Current Opinion in Genetics & Development*, vol. 48, 2018, pp. 121–127., doi:10.1016/j.gde.2017.11.007.

Peng, Yating, et al. “Circular RNA Profiling Reveals That circCOL3A1-859267 Regulate Type I Collagen Expression in Photoaged Human Dermal Fibroblasts.” *Biochemical and Biophysical Research Communications*, vol. 486, no. 2, 2017, pp. 277–284.,

doi:10.1016/j.bbrc.2017.03.028.

Pertea, Mihaela, and Steven L Salzberg. "Between a Chicken and a Grape: Estimating the Number of Human Genes." *Genome Biology*, vol. 11, no. 5, 2010, p. 206., doi:10.1186/gb-2010-11-5-206.

Raczynska, Katarzyna Dorota, et al. "Involvement of the Nuclear Cap-Binding Protein Complex in Alternative Splicing in *Arabidopsis Thaliana*." *Nucleic Acids Research*, vol. 38, no. 1, 2009, pp. 265–278., doi:10.1093/nar/gkp869.

Salzman, Julia, et al. "Cell-Type Specific Features of Circular RNA Expression." *PLoS Genetics*, vol. 9, no. 9, 2013, doi:10.1371/journal.pgen.1003777.

Salzman, Julia, et al. "Circular RNAs Are the Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types." *PLoS ONE*, vol. 7, no. 2, 2012, doi:10.1371/journal.pone.0030733.

Sanger, H L, et al. "Viroids Are Single-Stranded Covalently Closed Circular RNA Molecules Existing as Highly Base-Paired Rod-like Structures." *Proceedings of the National Academy of Sciences*, vol. 73, no. 11, 1976, pp. 3852–3856., doi:10.1073/pnas.73.11.3852.

Shen, Yudong, et al. "Identification and Characterization of Circular RNAs in Zebrafish." *FEBS Letters*, vol. 591, no. 1, 2016, pp. 213–220., doi:10.1002/1873-3468.12500.

Stumpf, M. P. H., et al. "Estimating the Size of the Human Interactome." *Proceedings of the National Academy of Sciences*, vol. 105, no. 19, 2008, pp. 6959–6964., doi:10.1073/pnas.0708078105.

Sun, Xiaoyong, et al. "Integrative Analysis of *Arabidopsis Thaliana* Transcriptomics Reveals Intuitive Splicing Mechanism for Circular RNA." *FEBS Letters*, vol. 590, no. 20, 2016, pp. 3510–3516., doi:10.1002/1873-3468.12440.

Suzuki, H. "Characterization of RNase R-Digested Cellular RNA Source That Consists of Lariat and Circular RNAs from Pre-mRNA Splicing." *Nucleic Acids Research*, vol. 34, no. 8, 2006, doi:10.1093/nar/gkl151.

Uhr, K., et al. "Association of MicroRNA-7 and Its Binding Partner CDR1-AS with the Prognosis and Prediction of 1st-Line Tamoxifen Therapy in Breast Cancer." *Scientific Reports*, vol. 8, no. 1, 2018,

doi:10.1038/s41598-018-27987-w.

Wang, Bing-Bing, and Volker Brendel. *Genome Biology*, vol. 5, no. 12, 2004, doi:10.1186/gb-2004-5-12-r102.

Wang, Peter L., et al. “Circular RNA Is Expressed across the Eukaryotic Tree of Life.” *PLoS ONE*, vol. 9, no. 3, 2014, doi:10.1371/journal.pone.0090859.

Xin, Ruijiao, et al. “IsoCirc Catalogs Full-Length Circular RNA Isoforms in Human Transcriptomes.” *Nature Communications*, vol. 12, no. 1, 2021, doi:10.1038/s41467-020-20459-8.

Yang, Yun, et al. “Extensive Translation of Circular RNAs Driven by N6-Methyladenosine.” *Cell Research*, vol. 27, no. 5, 2017, pp. 626–641., doi:10.1038/cr.2017.31.

Ye, Chu-Yu, et al. “Widespread Noncoding Circular RNAs in Plants.” *New Phytologist*, vol. 208, no. 1, 2015, pp. 88–95., doi:10.1111/nph.13585.

Ye, Jiazhen, et al. “AtCircDB: a Tissue-Specific Database For Arabidopsis circular RNAs.” *Briefings in Bioinformatics*, vol. 20, no. 1, 2017, pp. 58–65., doi:10.1093/bib/bbx089.

Zhang, Peijing, et al. “CircPlant: An Integrated Tool for CircRNA Detection and Functional Prediction in Plants.” *Genomics, Proteomics & Bioinformatics*, vol. 18, no. 3, 2020, pp. 352–358., doi:10.1016/j.gpb.2020.10.001.

Zhang, Pei, et al. “A Large-Scale Circular RNA Profiling Reveals Universal Molecular Mechanisms Responsive to Drought Stress in Maize and Arabidopsis.” *The Plant Journal*, vol. 98, no. 4, 2019, pp. 697–713., doi:10.1111/tpj.14267.

PUBLIKACJE WCHODZĄCE W SKŁAD ROZPRAWY
DOKTORSKIEJ

1. Philips A, **Nowis K**, Stelmaszczuk M, Jackowiak P, Podkowiński J, Handschuh L, Figlerowicz M;

Expression Landscape of circRNAs in Arabidopsis thaliana Seedlings and Adult Tissues

Frontiers in Plant Science (2020), <https://doi.org/10.3389/fpls.2020.576581>

Pięcioletni IF: 7,3, MEiN 100



Expression Landscape of circRNAs in *Arabidopsis thaliana* Seedlings and Adult Tissues

Anna Philips¹, Katarzyna Nowis¹, Michał Stelmaszczyk¹, Paulina Jackowiak¹, Jan Podkowiński¹, Luiza Handschuh¹ and Marek Figlerowicz^{1,2*}

¹ Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, ² Institute of Computing Science, Poznan University of Technology, Poznan, Poland

OPEN ACCESS

Edited by:

Mathew G Lewsey,
La Trobe University, Australia

Reviewed by:

Chengcheng Zhong,
Mineral Resources – CSIRO, Australia
Jianfei Zhao,
University of Pennsylvania,
United States

*Correspondence:

Marek Figlerowicz
marekf@ibch.poznan.pl

Specialty section:

This article was submitted to
Plant Cell Biology,
a section of the journal
Frontiers in Plant Science

Received: 26 June 2020

Accepted: 25 August 2020

Published: 10 September 2020

Citation:

Philips A, Nowis K, Stelmaszczyk M,
Jackowiak P, Podkowiński J,
Handschuh L and Figlerowicz M
(2020) Expression Landscape of
circRNAs in *Arabidopsis thaliana*
Seedlings and Adult Tissues.
Front. Plant Sci. 11:576581.
doi: 10.3389/fpls.2020.576581

RNA-seq is currently the only method that can provide a comprehensive landscape of circular RNA (circRNAs) in the whole organism and its particular organs. Recent years have brought an increasing number of RNA-seq-based reports on plant circRNAs. Notably, the picture they revealed is questionable and depends on the applied circRNA identification and quantification techniques. In consequence, little is known about the biogenesis and functions of circRNAs in plants. In this work, we tested two experimental and six bioinformatics procedures of circRNA analysis to determine the optimal approach for studying the profiles of circRNAs in *Arabidopsis thaliana*. Then using the optimized strategy, we determined the accumulation of circular and corresponding linear transcripts in plant seedlings and organs. We observed that only a small fraction of circRNAs was reproducibly generated. Among them, two groups of circRNAs were discovered: ubiquitous and organ-specific. The highest number of circRNAs with significantly increased accumulation in comparison to other organs/seedlings was found in roots. The circRNAs in seedlings, leaves and flowers originated mainly from genes involved in photosynthesis and the response to stimulus. The levels of circular and linear transcripts were not correlated. Although RNase R treatment enriches the analyzed RNA samples in circular transcripts, it may also have a negative impact on the stability of some of the circRNAs. We also showed that the normalization of NGS data by the library size is not proper for circRNAs quantification. Alternatively, we proposed four other normalization types whose accuracy was confirmed by ddPCR. Moreover, we provided a comprehensive characterization of circRNAs in *A. thaliana* organs and in seedlings. Our analyses revealed that plant circRNAs are formed in both stochastic and controlled processes. The latter are less frequent and likely engage circRNA-specific mechanisms. Only a few circRNAs were organ-specific. The lack of correlation between the accumulation of linear and circular transcripts indicated that their biogenesis depends on different mechanisms.

Keywords: circRNA, *Arabidopsis thaliana*, organ-specific, RNA-seq, droplet digital PCR, data normalization

INTRODUCTION

Circular RNAs (circRNAs) have recently emerged as a new, large class of noncoding RNAs with functions that, in most cases, remain unknown. To date, the abundance and evolutionary conservation of circRNAs have been shown across different species within the animal and plant kingdoms (Memczak et al., 2013; Salzman et al., 2013; Wang et al., 2014; Gao et al., 2015). However, more detailed studies of circRNAs thus far have only been performed in humans and several model animals [mainly cancer- and other disease-related studies (Qu et al., 2018)], leaving circRNAs in plants largely unexplored. CircRNAs have been identified in *Oryza sativa* (Lu et al., 2015; Ye et al., 2015), *Arabidopsis thaliana* (Liu et al., 2017), *Zea mays* (Chen et al., 2018), *Solanum lycopersicum* (Zuo et al., 2016), *Glycine max*, *Gossypium hirsutum* (Zhao et al., 2017), *Triticum aestivum* (Wang et al., 2016), *Hordeum vulgare* (Darbani et al., 2016), etc. A database listing circRNAs identified in plants has recently been updated based on publicly available resources (Chu et al., 2017). Unfortunately, due to substantial differences in the procedures applied in the particular experiments (lack of standards in the analyses of circRNA), the generated data cannot be directly compared. For example, the number of circRNAs identified in different plants ranges from 62 in *H. vulgare* (Darbani et al., 2016) to approximately 10,000 in *S. lycopersicum* (Tan et al., 2017). The same phenomenon is also observed if only one plant species is considered; e.g., a number of circRNAs found in *A. thaliana* vary between 6 (Wang et al., 2014) and 6,012 (Ye et al., 2015), depending on the experimental approach. According to PlantcircBase (<http://ibi.zju.edu.cn/plantcircbase/index.php>; 1 March 2019), the total number of circRNAs in *A. thaliana* identified in different studies reached 38,938.

A. thaliana is currently the best-characterized plant in terms of circRNAs. A large fraction of these molecules was identified based on RNA-seq data generated earlier to study gene expression levels. Thus, RNA isolation and sequencing protocols were rarely optimized for circRNA research. During bioinformatics analyses, it was routinely assumed that a circRNA is valid if its presence is supported by at least two so-called back-spliced reads (reads overlapping a circRNA splicing site). The depth of the library was normally not taken into consideration, and thus, the numbers of circRNA candidates varied markedly between studies. Accordingly, questions about the levels of circRNA accumulation in plants and, consequently, their functional relevance have not been satisfactorily answered to date. Some reports have suggested that circRNAs in *A. thaliana* are expressed in tissue and developmental stage-specific manners (Sun et al., 2016; Chen et al., 2017; Liu et al., 2017). Additionally, Pan and coworkers (Pan et al., 2018) identified 1,583 heat-stress specific circRNAs that were absent in the control samples. In only a few cases, selected circRNAs identified based on RNA-seq data were experimentally validated by quantitative PCR (qPCR) (Wang et al., 2014; Liu et al., 2017; Cheng et al., 2018; Pan et al., 2018). Unfortunately, the obtained results were not used to optimize RNA-seq data normalization.

Critical review of the reports on the identification and differential accumulation of circRNAs in plants revealed that

the published results are largely inconsistent due to the lack of standardized methods. To address this issue, we tested two major protocols of circRNA-seq library preparation (with and without RNase R treatment) and several approaches of RNA-seq data normalization that are routinely used to quantify other transcripts. We showed that the treatment of RNA samples with RNase R significantly increased the number of identified circRNAs. However, the spectrum of highly abundant circRNAs was reduced. Moreover, to compare the levels of circRNA accumulation in various plant organs, it is necessary to apply a circRNA-dedicated method of RNA-seq library normalization. After testing six normalization methods, we identified those that provided data that were most coherent with droplet digital PCR (ddPCR) results. We also noticed that the best correlation was observed when comparing samples collected from the same organs.

Next, using the optimized strategy, we investigated the accumulation of circular and corresponding linear transcripts in *A. thaliana* seedlings and organs, which led to the identification of ubiquitous and organ-specific circRNAs. Our results revealed that, in many cases, the abundance of circRNAs was not directly linked to the level of parent gene expression, which suggested the existence of circRNA-dedicated regulatory mechanisms.

MATERIALS AND METHODS

Plant Material and RNA Extraction

Seeds of wild-type (Col-0) *A. thaliana* plants (deposition number: CS70000) were obtained from Arabidopsis Biological Resource Center (ABRC). The formal identification of the plant material used in the study was performed by J. Ecker laboratory (plant material donor) using Illumina sequencing by synthesis technology. Seeds were sterilized by placing them in filter tubes and washing them with 70% ethanol, distilled water, 0.01% Amistar 250 SC (Syngenta) and distilled water, using a vacuum pump. Next, the seeds were exposed to stratification for 4 days at 4°C in 0.1% agarose solution. Plants were grown using Arasystem (Betatech) and Jiffy-7 peat pellets in a growth chamber with 16 h of light at 23°C and 8 h of dark at 18°C. Whole seedlings at stage 1.0 (cotyledons fully open), whole roots and leaves at stage 3.90 (rosette growth complete) and whole flowers at stage 6.90 (flowering complete, flowers are no longer produced) (Boyes et al., 2001) were collected, frozen in liquid nitrogen and stored at -80°C until RNA isolation. Total RNA was isolated from 100 mg of powdered plant samples using the mirVana miRNA Isolation Kit (Thermo Fisher Scientific). For each replicate, 50 seedlings or 30 plants were used to obtain powdered samples. Subsequently, 10 µg of total RNA was treated with 2 U of Turbo DNase (Thermo Fisher Scientific) and purified with the QIAquick Nucleotide Removal Kit (Qiagen). RNA quality and integrity were determined using capillary electrophoresis (2100 Bioanalyzer, Agilent) with the Plant RNA Nano Assay. For further analyses, we used samples with the RNA integrity number (RIN) >9 for roots and flowers and >7.5 for tissues with a high content of chloroplast rRNA (seedlings and leaves).

Library Preparation and Sequencing

Total RNA samples (isolated as described above) were divided into two portions: one portion was used to prepare an RNA-seq library without prior RNase R treatment (R-), and the other portion was used to prepare an RNA-seq library from RNase R-treated RNA (R+). For library preparation, we used 2.5 µg of total RNA. According to the first protocol (R-), 2.5 µg of total RNA was treated with Ribo-Zero (Illumina) following the manufacturer's recommendation. The Ribo-Zero rRNA Removal Kit (Plant Leaf) was used for seedlings, leaves and flowers, and the Ribo-Zero rRNA Removal Kit (Plant Seed/Root) was used for roots. The level of rRNA depletion was determined by capillary electrophoresis (2100 Bioanalyzer, Agilent) with the Plant RNA Pico Assay.

The second protocol (R+) also included treatment of 2.5 µg of total RNA with Ribo-Zero. Then, 500 ng of the Ribo-Zero-treated RNA was denatured at 70°C for 3 min, treated with 10 U of RNase R (Epicentre) at 37°C for 1 h and subsequently purified with the QIAquick Nucleotide Removal Kit (Qiagen). The quality and integrity of the resultant RNA was tested using capillary electrophoresis (2100 Bioanalyzer, Agilent) with the Plant RNA Pico Assay.

One hundred nanogram aliquots of R- and R+ RNA samples from seedlings, roots, leaves, and flowers were prepared in four biological replicates. Next, 8 indexed libraries per biological replicate (16 in total) were prepared separately using the TruSeq RNA Sample Preparation Kit (Illumina) in accordance with the manufacturer's instructions with minor changes. The protocol was modified by omitting the mRNA purification step and reducing reaction volumes by one-third at subsequent steps. The libraries were submitted to qualitative analysis using capillary electrophoresis (2100 Bioanalyzer, Agilent) with the High Sensitivity DNA Assay and quantitative analysis with a Qubit fluorometer (Invitrogen) prior to sequencing. The samples were sequenced with the Genome Analyzer IIx (Illumina) and 108-bp, paired-end protocol.

CircRNA Identification

Raw reads were trimmed and quality-filtered with AdapterRemoval [version 1.5.4; (Schubert et al., 2016)]. As a result, remnant adapter sequences were removed, and read ends of low-quality were trimmed to save bases at the 3' and 5' end with a minimum quality score of 30 (99.9% base call accuracy). Reads shorter than 20 bases were removed. The rRNA content was assessed with SortMeRNA (version 2.1 [Kopylova et al., 2012]).

CircRNAs were identified using the protocol recommended by (Gao et al., 2015). Briefly, filtered reads were mapped to the reference genome (TAIR10) with BWA mem [version 0.7.10; (Li and Durbin, 2009)]. Subsequently, we used CIRI2 [version v2.0.5; (Gao et al., 2017)] and find_circ (Memczak et al., 2013) to identify the reads overlapping back-splicing sites. As a result of CIRI2, we obtained a list of circRNA candidates that were in summary supported by at least 2 back-spliced reads. CircRNAs not identified with find_circ were filtered out from final results.

For the selected circRNAs, their formation was confirmed by PCR. For each individual circRNA, two PCRs were performed:

the first reaction involved divergent primers (circular transcript specific), and the second reaction involved convergent primers (linear transcript specific). As a control, analogous PCRs with gDNA as a template were performed. All primers are listed in **Additional file 1: Supplementary Table 1**. Prior to PCR, cDNA was obtained in the reverse transcription (RT) reaction, including 1 µg of Turbo DNase-treated total RNA and iScript Reverse Transcription Supermix (Bio-Rad). PCR mixtures (50 µl of the total volume) contained 1 µl of cDNA template from the 20-µl RT reaction or 10 ng of gDNA, 1.25 U of Taq DNA polymerase (5 U/µl), 1.75 mM MgCl₂, 0.2 mM dNTPs, 1× Taq Buffer with (NH₄)₂SO₄ (Thermo Fisher Scientific), and 0.3 µM of each primer (Genomed). PCR was performed in a T-100 thermal cycler (Bio-Rad) using the following program: preheating at 95°C for 3 min, followed by 40 cycles (with divergent primers) or 35 cycles (with convergent primers) at 95°C for 30 s, 59°C to 62°C (dependent on the melting temperature (T_m) of the primers used – listed in **Additional file 1: Supplementary Table 1**) for 30 s and 72°C for 30 s. Finally, the reaction mixtures were subjected to elongation at 72°C for 5 min. PCR products were purified using the QIAquick PCR Purification Kit (Qiagen), resolved in 2% agarose gel for 70 min at 120 V in a Wide Mini-Sub Cell GT System (Bio-Rad) and poststained with 0.5 µg/ml EtBr solution for 20 min on a rocker. Gel images were obtained with a ChemiDoc XRS+ System (Bio-Rad) and analyzed with Image Lab Software (Bio-Rad). Purified PCR products were cloned using the TOPO TA Cloning Kit (Thermo Fisher Scientific) in accordance with the manufacturer's instructions. Clones were screened by EcoRI digestion (Thermo Fisher Scientific), and positive clones were purified using the QIAquick PCR Purification Kit (Qiagen) and sequenced using the 3130XL Genetic Analyzer (Applied Biosystem). The sequencing results were analyzed using SnapGene software (GSL Biotech LLC).

Quantitative Analysis of circRNA

The quantity of each circRNA was estimated based on the normalized back-spliced read count. Six normalization methods were tested: by the library size, by the sum of back-spliced reads, by the DESeq2 size factor, by the library size without rRNA and chloroplast RNA reads, by the sum of back-spliced reads without rRNA and chloroplast RNA reads and by the number of raw reads mapping to the endogenous marker *ACT2*. To determine the raw read numbers mapping to the selected gene, we applied the protocol proposed by (Peretea et al., 2016): filtered reads were mapped with HISAT2 (version 2.1.0; [Kim et al., 2015]) to the reference *A. thaliana* genome (TAIR10). Gene quantification analyses were performed with StringTie [version 1.3.3b; (Peretea et al., 2015)].

The distribution of circRNAs in organs and seedlings and the circRNA content in biological replicates were visualized using the VennDiagram R package. Heatmaps and pairwise comparison scatterplots were produced using the ggplot2 R package. The circRNA accumulation pattern in organs/seedlings on heatmaps was visualized with 4 discrete classes (high, medium, low, not detected). CircRNAs were grouped by hierarchical agglomerative clustering and divided between classes as follows: high for 5% of circRNAs with the highest accumulation in organs/seedlings;

medium - circRNAs with higher accumulation than 80% of circRNAs and lower than 5% of circRNAs with the highest accumulation; low - remaining circRNAs; not detected - circRNAs not detected in particular organ/seedlings. Linear counterparts were ordered corresponding to circRNAs. To evaluate associations between two variables (circRNAs and their linear counterparts; accumulation of circRNA obtained from RNA-seq and ddPCR) we calculated Pearson correlation using R package. GO enrichment analyses were conducted with agriGO (Tian et al., 2017).

ddPCR reactions were performed with the QX200 ddPCR System (Bio-Rad) in accordance with the manufacturer's protocol. ddPCR reactions were performed with two sets of primers: a pair of divergent primers or a pair of convergent primers. All primers used are listed in **Additional file 1: Supplementary Table 1**. Assay mixtures contained 1 μ l of cDNA template (obtained as described above), 10 μ l of QX200 ddPCR EvaGreen Supermix (Bio-Rad), and 0.1 μ M of each primer (Genomed). Thermal cycling conditions in the T-100 thermal cycler (Bio-Rad) were as follows: 95°C for 5 min, followed by 40 cycles of 95°C for 30 s and 59°C to 62°C (dependent on the T_m of the primers used - listed in **Additional file 1: Supplementary Table 1**) for 1 min, a final inactivation step at 4°C for 5 min followed by 90°C for 5 min and an infinite hold at 12°C. All steps were performed at a ramp rate of 2°C/s. The read-out was performed with the QX200 Droplet Reader (Bio-Rad). The results were analyzed using the QuantaSoft Analysis Pro software (Bio-Rad). Each experiment was carried out in three independent biological replicates. To exclude variances in the RT reaction between different templates, for each biological replicate, the *ACT2* expression level was determined in parallel in four technical repeats as a reference for normalization.

RESULTS

CircRNA Identification in *A. thaliana* Seedlings and Organs

Considering the problems encountered by other researchers studying circRNA in plants, we attempted to evaluate and optimize the circRNA-dedicated protocols for RNA-seq and sequencing data analysis. Our efforts were focused on RNA-seq because it is currently the only method that can provide a comprehensive landscape of circRNAs in the whole plant and its particular organs. In the first step, we isolated total RNA from seedlings, roots, leaves, and flowers of wild-type *A. thaliana* (Columbia ecotype, abbreviated Col-0) and prepared RNA-seq libraries using two protocols that are widely applied in circRNA studies. In line with the first approach, we used total RNA depleted of rRNA (samples contained both linear and circular transcripts and resultant libraries were denoted with R-, see **Additional file 1: Supplementary Table 2A**; samples from leaves and seedlings contained also traces of chloroplast rRNA, see **Additional file 2: Supplementary Figure 1**). In accordance with the second approach, we used total RNA depleted of rRNA and subsequently treated with RNase R (samples contained circRNA

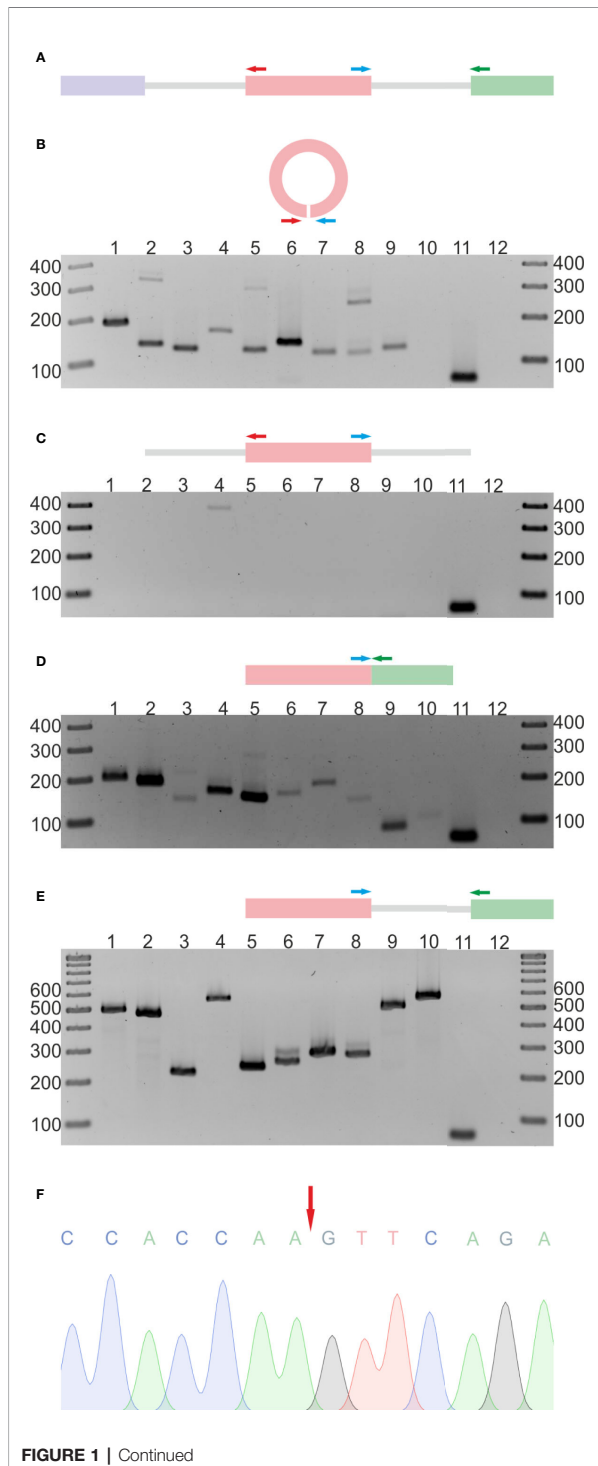
and traces of linear RNA, and resultant libraries were designated with R+, see **Additional file 1: Supplementary Table 2B**). All RNA-seq experiments were performed in 4 biological replicates.

On average 60,292,166 paired-end reads were obtained for R- libraries; 99.54% of the reads passed trimming and quality filtration, and 95.30% of them mapped to the *A. thaliana* reference genome (TAIR10). For the R+ libraries, 25,932,926 paired-end reads, on average, were obtained; 99.00% of the reads passed trimming and quality filtration, and 91.87% of them mapped to the *A. thaliana* reference genome (TAIR10).

For the R- libraries, from 0.6% to 7.57% of the reads mapped to the rRNA sequences (roots and leaves, respectively). For the R+ libraries, the percentage of reads that mapped to the rRNA sequences varied from 0.27% in roots to 3.18% in flowers. This observation was consistent with the capillary electrophoresis results (**Additional file 2: Supplementary Figure 1**).

For circRNA discovery, we initially used CIRI2 (Gao et al., 2015) as the best-performing *de novo* identifying program (Szabo and Salzman, 2016; Hansen, 2018). In total, we detected 5,235 and 15,253 distinct circRNA candidates in R- and R+, respectively, supported by at least two back-spliced reads (**Additional file 1: Supplementary Tables 3A, B**). A large fraction of the detected circRNA candidates was supported by 2 to 5 back-spliced reads (4,297 circRNAs = 82.1% in R- and 12,155 circRNAs = 79.6% in R+). However, we also identified 39 (R-) and 88 (R+) circRNAs supported by more than 100 back-spliced reads. Approximately half of the identified circRNAs had been reported previously, 37.7% and 24.8% [AtCircDB (Ye et al., 2017)], and 42.5% and 28.9% [PlantcircBase, (Chu et al., 2017)] in R- and R+, respectively. In general, the majority of circRNAs (R-: 88.2%, R+: 97.3%) were derived from annotated genes. Considering that we only found 4,619 (R-) and 14,848 (R+) genes producing circRNAs, the median number of circRNAs per gene was 1. However, for 2 genes, more than 10 circRNAs were found in both libraries. A circRNA typically overlapped 1 to 4 exons (1 exon 31.5% and 28.3%; 2 exons 32.8% and 32.9%; 3 exons 15.9% and 19.8%; and 4 exons 8.6% and 9.4% in R- and R+, respectively), but we also identified 24 (R-) and 36 (R+) circRNA candidates enclosing more than 10 exons.

In the next step, the RNA-seq results were validated by reverse transcription PCR (RT-PCR). We selected 10 circRNA candidates that were identified in both the R- and R+ libraries. Nine of the candidates were found in leaves, and 1 was absent in leaves but present in 2 other organs and in seedlings (**Additional file 1: Supplementary Tables 3A, B**). As a template in the RT-PCR reaction, we used total RNA from leaves (neither depleted of rRNA nor treated with RNase R). Regions of circRNAs encompassing back-splice junctions were amplified using circRNA-specific divergent primers (each pair specific for an individual circRNA, for details, see **Additional file 1: Supplementary Table 1, Figure 1A**). In nine cases, RT-PCR products of the predicted lengths were obtained (**Figure 1B**). The product was not observed, as expected, for the circRNA that, according to the RNA-seq data, was absent in leaves (lane 10). Three randomly selected products were cloned and sequenced, which confirmed the presence of back-splice junctions (**Figure 1F**,



Additional file 2: Supplementary Figures 3A, B, Supplementary Figure 4B). To rule out the possibility of genomic rearrangements, we performed PCR involving the same set of divergent primers and

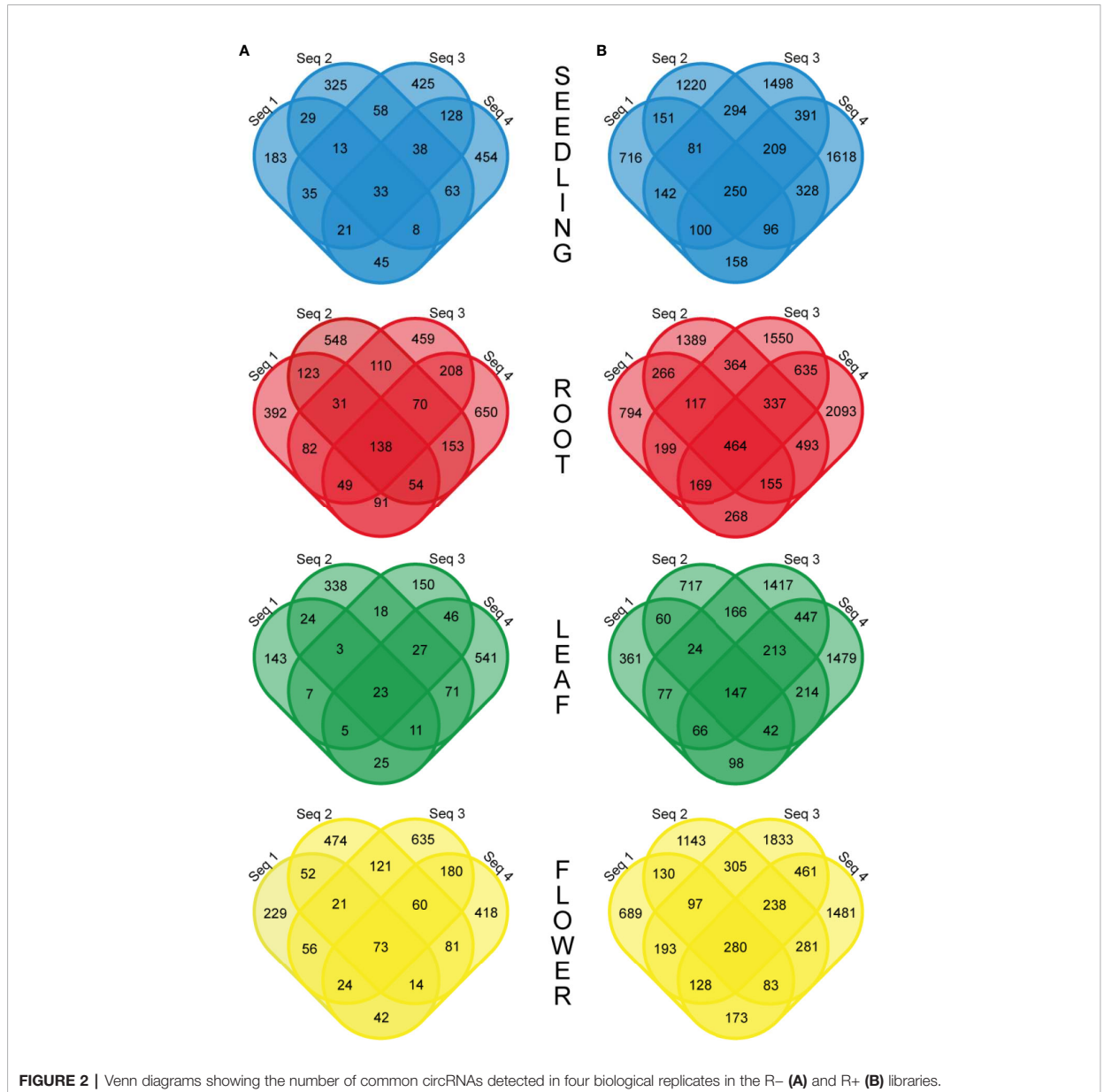
FIGURE 1 | Validation of 10 selected circRNAs identified in *A. thaliana*. (A) Schematic representation of the PCR strategy. Colored rectangles and gray lines represent exons and introns, respectively. Divergent (red and blue arrows) and convergent (blue and green arrows) primers were used to detect circular RNAs and their linear counterparts. (B) Detection of circRNAs by RT-PCR with leaf cDNA as template and divergent primers. Lanes 1 to 10 show PCR products corresponding to circRNA candidates. Lanes 11 and 12 represent positive (leaf cDNA template) and negative (no template) controls, respectively, with convergent primers for *ACT2*. (C) PCR products obtained with leaf gDNA and divergent primers. Lanes 1 to 10 show PCR products corresponding to circRNA candidates. Lanes 11 and 12 represent positive (leaf cDNA template) and negative (no template) controls, respectively, with convergent primers for *ACT2*. (D) Detection of linear transcripts by RT-PCR with leaf cDNA as template and convergent primers. Lanes 1 to 10 show PCR products corresponding to linear counterparts of circRNA candidates. Lanes 11 and 12 represent positive (leaf cDNA template) and negative (no template) controls, respectively, with convergent primers for *ACT2*. (E) PCRs with leaf cDNA and convergent primers. Lanes 1 to 10 show PCR products corresponding to linear counterparts of circRNA candidates. Lanes 11 and 12 represent positive (leaf cDNA template) and negative (no template) controls, respectively, with convergent primers for *ACT2*. (F) Sequencing chromatogram of the circRNA 1 (5:1106879-1107381) amplification product. The red arrow indicates the back-splice site. The uncropped gel images are shown in **Additional file 2: Supplementary Figure 2**. Additional Sanger sequencing results are shown in **Additional file 2: Supplementary Figure 3**. CircRNA IDs: 1) 5:1106879-1107381; 2) 1:382803-383130; 3) 2:19672192-19672380; 4) 1:30349032-30349237; 5) 2:14529841-14530023; 6) 4:13026572-13027129; 7) 3:3172073-3172253; 8) 2:16614068-16615133; 9) 2:17272368-17272938; 10) 1:3674466-3675958.

genomic DNA (gDNA) as a template (Figure 1C). No product was obtained except in one reaction (lane 4). Cloning and sequencing of this product showed that it resulted from mispriming to the AT5G37790 locus (Supplementary Figure 3C). In addition, to confirm the formation of linear counterparts of all circRNAs and to show that the gDNA contained the expected arrangement of exons, we performed PCR using, respectively, complementary DNA (cDNA) (Figure 1D) and gDNA (Figure 1E) as templates and convergent primers (each pair specific for the individual transcript/gene) flanking the canonical splice junctions located downstream of the circRNA-producing exon (Additional file 1: Supplementary Table 1, Figure 1A). As controls, we performed PCR with convergent primers specific for *ACTIN2* (*ACT2*, AT3G18780). In the positive control (lane 11), leaves cDNA were used as a template, and in the negative (lane 12), no template was added. The obtained results were fully consistent with the RNA-seq data obtained for both the R- and R+ libraries.

More detailed investigations of the RNA-seq dataset revealed that the fraction of circRNAs identified in four biological replicates was surprisingly small and varied from 1.3% (R-) to 2.7% (R+) for the leaf libraries and from 4.4% (R-) to 5.0% (R+) for the root samples (Figure 2). Interestingly, some circRNAs supported by back-spliced reads in the R- libraries were not present in the R+ libraries (Additional file 1: Supplementary Table 3C). To confirm their presence, we selected 7 circRNAs identified by RNA-seq analysis only in R- library. Five of the circRNAs were identified in R- and not identified in R+ libraries from leaves and 2 were absent in leaves but identified in R- libraries from seedlings and roots (Additional file 1: Supplementary Table 3C). In addition, we selected a circRNA

candidate derived from the chloroplast chromosome that was found in both R- and R+ libraries from leaves (**Additional file 1: Supplementary Tables 3A, B**). Then, we used standard RT-PCR to determine whether they were present in total RNA extracted from leaves. In the PCR, we used cDNA (obtained as described in Materials and Methods) and divergent primers specific for individual circRNAs (**Additional file 1: Supplementary Table 1**). As in the former experiment, primers specific for *ACT2* were used in control reactions: positive included cDNA from leaves and negative contained no template. In 7 reactions, the products of predicted lengths were obtained (**Additional file 2:**

Supplementary Figure 4A, lanes 1–7). Surprisingly, the circRNA identified only in roots was amplified (**Additional file 2: Supplementary Figure 4A**, lane 4). Nevertheless, the product for circRNA identified only in seedlings was not formed (**Additional file 2: Supplementary Figure 4A**, lane 8). PCR product from lane 3 was cloned and sequenced, which confirmed the presence of back-splice junction (**Additional file 2: Supplementary Figure 4B**). These results provided an additional piece of evidence that circRNAs not found in R+ libraries were present *in planta*. Additionally, in this experiment, we confirmed the formation of the circRNA candidate derived



from the chloroplast chromosome that was found in both R– and R+ libraries from leaves (**Additional file 2: Supplementary Figure 4A**, lane 7).

The above results showed that a large fraction of circRNAs did not form reproducibly in plants. Moreover, some circRNAs with considerable functional potential, i.e., accumulating in plants to high levels, were not identified in R+ libraries. Consequently, in further analyses, we used RNA-seq data obtained with R– libraries, as we intended to study highly abundant, reproducible circRNAs with functional potential.

Quantitative Analysis of circRNA in Seedlings and Plant Organs

In former reports on circRNAs, their accumulations in whole plants and in different organs were usually determined based on RNA-seq data (read counts) without further validation by other experimental methods. To obtain detailed insight into this issue, we compared the results of RNA-seq-based quantitative analysis of circRNAs with corresponding data obtained using ddPCR. To achieve this goal, we selected 8 circRNAs and determined their levels in seedlings, roots, leaves, and flowers by ddPCR. The summarized ddPCR results based on 3 biological replicates are shown in **Table 1** col. 1 (for details see **Additional file 1: Supplementary Table 4**). These results were used as a reference for determining the best method of RNA-seq data normalization. First, the back-spliced read counts obtained for the individual circRNA were normalized using the total number of reads in a library (**Table 1**, col. 2), as this type of normalization is the most common in circRNA studies (Ye et al., 2015; Wang et al., 2016; Liu et al., 2017). The result of Pearson correlation coefficient calculated for this dataset was 0.67. Thus, the normalization by the library size did not provide satisfying results. Next, the read counts were normalized using the total number of back-spliced reads in the library (**Table 1**, col. 3). This approach provided better results, with the Pearson correlation coefficient reaching 0.77. Finally, the read counts were normalized using the size factors calculated by the DESeq2 program (**Table 1**, col. 4). This type of normalization is frequently used to analyze differential gene expression (see Materials and Methods). The obtained correlation coefficient was 0.76.

The difficulties with proper RNA-seq data normalization motivated us to compare the content of the RNA-seq libraries. We found significant differences in the composition of organ/seedling transcriptomes. In particular, the numbers of reads mapping to rRNA genes and the chloroplast genome were markedly distinct (**Additional file 1: Supplementary Table 2**). This phenomenon is the most likely explanation for the ineffective normalization based on the library size. Consequently, we excluded from the libraries the reads from rRNA and chloroplast RNA. We normalized this new set of RNA-seq data and again compared them with the results of the ddPCR-based analysis (**Table 1**, col. 5 and 6). The correlation coefficient reached 0.76. Thus, the transcriptome composition is an important factor affecting the results of quantitative RNA-seq-based analysis of circRNA. However, rRNA and

chloroplast-derived reads were not solely responsible for the differences.

Similar results were obtained when we used reference genes as an internal, steady factor to normalize the RNA-seq data. The best results were obtained for *ACT2* gene (**Table 1**, col. 7), which is known to be expressed at a relatively stable level throughout the whole plant (Dheda et al., 2004; Radonic et al., 2004). In this case, the Pearson correlation coefficient was 0.77, (**Table 1**, col. 3). It is noteworthy that the Pearson correlation coefficient between the ddPCR and normalized RNA-seq results was high if a single organ was considered regardless of the normalization type applied. The correlation values ranged from 0.88 (flowers) to 0.90 (roots) and 0.92 (leaves). The correlation coefficient was relatively lower (0.76) only for seedlings, which might be explained by the generally low abundances of circRNAs in the developing plant (**Table 1**).

The results obtained indicated that four of tested normalization methods, namely, normalization by: (i) the number of back-spliced reads, (ii) the DESeq2 size factor (iii) the library size without rRNA and chloroplast RNA reads, (iii) the number of back-spliced reads without rRNA and chloroplast RNA reads and (iv) housekeeping gene *ACT2*, reached a similar Pearson correlation coefficient between the ddPCR and RNA-seq results of 0.76 to 0.77 and might be used alternatively.

Moreover, analysis of the ddPCR results showed that 50 copies of circRNA per microgram of total RNA (cp/μg) represented the minimal value for quantification of circRNA using this method. Below this threshold, the standard deviation and average values were comparable, which made the results unreliable (**Additional file 1: Supplementary Table 4**).

To examine circRNA differential accumulation using RNA-seq data, we applied normalization by the library size without reads that mapped to the chloroplast genome and rRNA genes. Our analysis included only those circRNAs that were represented by at least 10 RNA-seq normalized reads (average of four biological replicates), which corresponded to the threshold of 50 cp/μg established by the ddPCR and were additionally identified by find_circ program (**Additional file 1: Supplementary Table 5**). One hundred twenty seven circRNA candidates met this criterion; 109, 87, 79 and 99 circRNAs in seedlings, roots, leaves, and flowers, respectively. One hundred twelve (88.2%) and 67 (52.8%) of them were previously identified and deposited in online databases (PlantcircBase and AtCircDB respectively). Fifty two (40.9%) circRNAs were present in all 3 organs and in seedlings. The largest number of unique circRNAs (8 with and 6 without chloroplast circRNAs) was found in roots (**Figures 3A, B**). CircRNAs derived from the chloroplast genome were present in seedlings (24), roots (2), leaves (12) and flowers (14). Only 7 of the circRNAs were common to all organs and seedlings. Five chloroplast circRNAs were seedling-specific, 2 leaf-specific, whereas there were no root- and flower-specific ones (**Figure 3C**).

A. *thaliana* Genes Giving Rise to circRNAs

Gene ontology (GO) analysis of chloroplast genes giving rise to circRNAs did not reveal any significant enrichment. Consecutive GO analysis revealed that the circRNAs in seedlings, leaves and

TABLE 1 | Abundances of circRNAs in seedlings and plant organs determined by ddPCR and RNA-seq with different types of data normalization (for the extended version of this table, see **Additional file 1: Supplementary Table 4**).

Column number Sample	1 ddPCR	2	3	4	5 RNA-seq normalized reads	6	7			
circRNA name	cp/μg	st. dev.	Library size	Back-spliced reads	DESeq2 size factor	Library size—without rRNA and chloroplast reads	Back-spliced reads—without rRNA and chloroplast reads			
circRNA ID	Organ						ACT2 (AT3G18780)			
A	4:4196362-4197112	Seedling	70.9	42.5	15.5	299028.5	1256309309.2	34.1	416842.1	32441.5
B	5:1106879-1107381	Seedling	75.9	46.3	32.5	639122.0	2401751918.5	56.9	846227.8	62709.9
C	1:382803-383130	Seedling	6.3	5.6	4.6	99458.9	392892148.2	8.1	139727.3	10254.1
D	2:19672192-19672380	Seedling	8.1	2.3	3.8	75095.2	350726157.6	7.8	103507.7	9207.0
E	1:30349032-30349237	Seedling	6.6	7.3	4.2	85267.6	325177084.7	8.2	117946.2	8406.7
F	2:14529841-14530023	Seedling	4.4	3.9	1.8	32451.1	107648015.3	3.5	41682.9	2756.1
G	3:3172073-3172253	Seedling	8.6	7.7	1.0	17127.8	86565020.0	2.7	23909.8	2232.6
H	2:16614068-16615133	Seedling	57.5	36.6	2.2	42413.1	152047044.7	3.9	56545.5	3941.6
A	4:4196362-4197112	Root	310.8	58.0	62.8	514242.3	1459154931.5	64.3	515369.2	40115.6
B	5:1106879-1107381	Root	403.9	35.5	116.7	984781.5	2670437284.5	119.4	986902.9	73977.9
C	1:382803-383130	Root	50.8	8.1	19.8	169589.9	442625591.2	20.2	169930.3	12169.8
D	2:19672192-19672380	Root	42.0	8.0	15.8	136481.1	358654044.8	16.2	136774.5	10013.7
E	1:30349032-30349237	Root	8.5	3.7	15.1	136987.2	328463667.3	15.3	137248.5	9118.7
F	2:14529841-14530023	Root	1.7	3.0	0.4	2450.5	7547596.9	0.4	2458.2	223.8
G	3:3172073-3172253	Root	6.7	0.6	4.1	35013.7	84508071.1	4.2	35082.9	2362.5
H	2:16614068-16615133	Root	165.7	45.2	1.4	11901.1	30190387.4	1.5	11931.7	895.0
A	4:4196362-4197112	Leaf	712.3	103.7	30.1	1260098.8	4104055169.7	58.6	1358140.6	127464.0
B	5:1106879-1107381	Leaf	751.1	177.7	33.8	1561949.7	4696612483.6	70.4	1650893.6	145891.2
C	1:382803-383130	Leaf	87.5	23.4	5.5	189330.0	568864974.2	8.8	217091.3	17681.0
D	2:19672192-19672380	Leaf	43.8	9.6	2.9	121121.1	355550976.4	5.4	133771.1	11041.2
E	1:30349032-30349237	Leaf	14.2	12.4	3.2	107435.2	392817341.8	5.3	122590.9	12181.3
F	2:14529841-14530023	Leaf	21.5	18.4	1.9	86108.0	186248886.7	3.5	94131.7	5776.3
G	3:3172073-3172253	Leaf	13.2	5.6	0.8	35715.0	60958705.1	1.3	40667.9	1886.3
H	2:16614068-16615133	Leaf	413.2	111.2	3.1	151191.6	399562884.5	6.4	159512.9	12416.1
A	4:4196362-4197112	Flower	966.3	186.2	50.6	518142.0	1815139368.8	65.4	547677.8	52437.1
B	5:1106879-1107381	Flower	1095.2	233.1	104.2	1150954.9	3452046537.0	129.6	1215016.1	99180.1
C	1:382803-383130	Flower	128.6	48.6	11.7	127251.5	403191198.6	14.8	135097.2	11576.1
D	2:19672192-19672380	Flower	42.0	17.0	6.7	79017.3	206880247.3	8.0	83020.7	5931.3
E	1:30349032-30349237	Flower	16.0	9.0	11.6	141156.9	382052606.6	13.5	149205.6	11003.1
F	2:14529841-14530023	Flower	5.6	9.7	5.1	61782.3	155535668.2	6.1	65279.9	4431.7
G	3:3172073-3172253	Flower	12.2	7.2	4.0	42481.2	144215416.7	5.1	44790.1	4178.8
H	2:16614068-16615133	Flower	581.3	96.7	9.3	107236.4	329206126.5	11.1	112814.1	9570.6
Correlation ddPCR vs RNA-seq										
All samples			0.67	0.77	0.76	0.76	0.76	0.76	0.75	0.77
Seedling			0.75	0.75	0.75	0.76	0.76	0.78	0.75	0.76
Root			0.90	0.89	0.90	0.90	0.90	0.90	0.90	0.90
Leaf			0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
Flower			0.87	0.85	0.85	0.88	0.85	0.87	0.85	0.88

flowers predominantly came from genes involved in photosynthesis and the response to stimulus (seedling and flower) (p-value: <0.01), **Additional file 1: Supplementary Tables 6A–C** respectively. Statistically significant GO enrichment was not observed for genes producing circRNAs in roots.

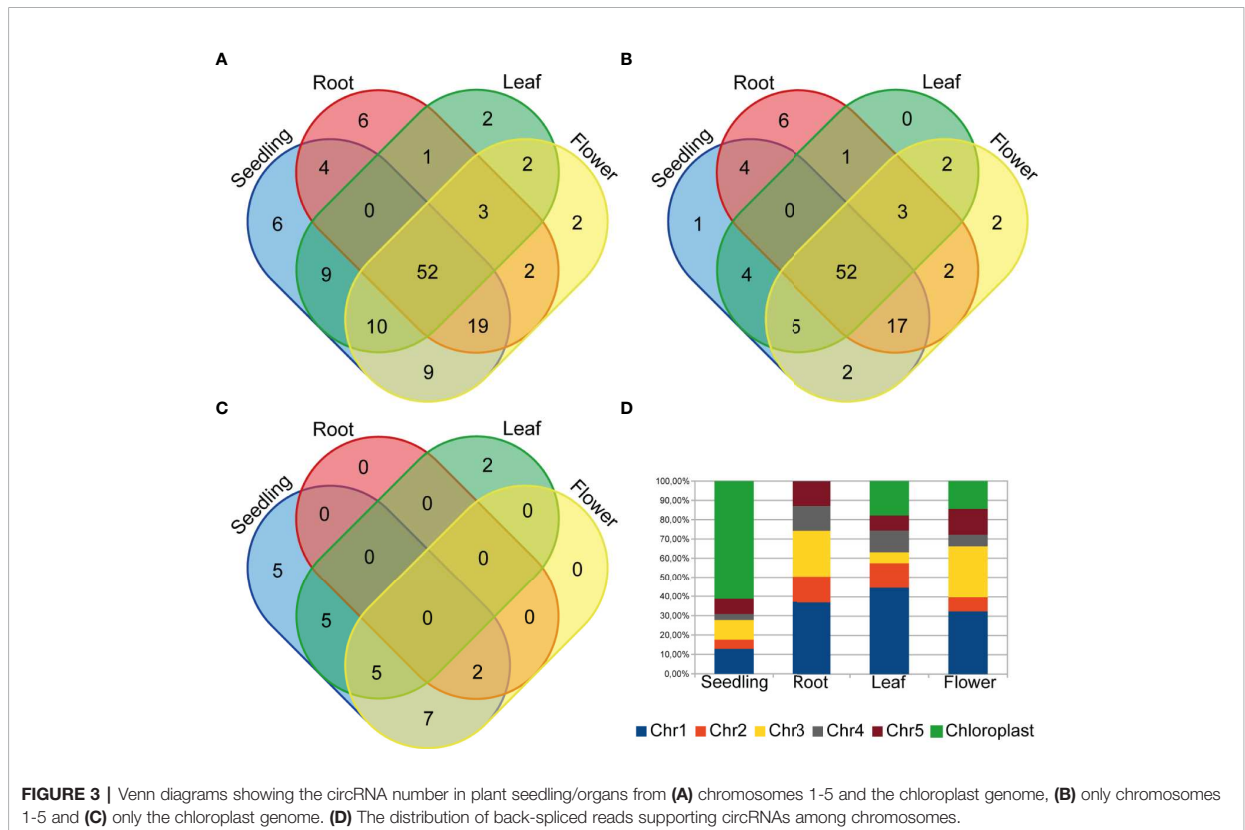
The analysis of back-spliced reads revealed the predominance of reads corresponding to chloroplast circRNAs in seedlings (**Figure 3D**). We also attempted to determine if any *A. thaliana* chromosomes were a preferable source of circRNAs. However, this analysis did not reveal any statistically significant differences.

Further analyses were focused on circRNAs that were present in at least two organs or in one organ and seedlings. A pairwise comparison revealed an organ-specific pattern of circRNA accumulation (**Figure 4A**, **Additional file 1: Supplementary Table 5**). The largest number of circRNAs with significantly increased accumulation (p-value ≤ 0.05) in comparison to other organs/seedlings was found in roots (16, 25 and 9 circRNA vs seedlings, leaves, and flowers, respectively, **Figures 4A, B**). Cellular process and organelle part GO terms (p-value <0.000001) were the only categories characteristic for the group of genes encoding differentially expressed circRNAs in leaves versus flowers (24, **Additional file 1: Supplementary Table 6D**). No other statistically significant GO enrichment was observed. The smallest number of circRNAs with increased accumulation was

identified in seedlings (1, 3 and 1 for seedlings vs roots, leaves, and flowers respectively, **Figure 4B**).

Accumulation of circRNAs and Their Linear Counterparts

Finally, we studied the correlation between the abundance of circRNAs and their linear counterparts. The expression profiles of circular and linear transcripts in different organs/seedlings arranged in a matching order are shown in **Figures 5A, B**, respectively. The Pearson correlation coefficient between the accumulation of corresponding circular and linear RNAs was 0.46 (0.52, 0.04, 0.09, and 0.36 in seedlings, roots, leaves, and flowers, respectively). Thus, in most cases, there was no relationship between the levels of circRNA and linear counterpart production. For 6 circRNAs, their accumulation levels were at least 1.5 times higher than those observed for the corresponding linear transcripts (in at least one organ/seedling; **Additional file 1: Supplementary Table 5**). Four originated from intergenic regions, and 2 (1:12195268-12196524, 3:17169794-17170903) were from transcripts of genes AT1G33615 and AT3G46614 (encodes ncRNA). Abundance of circRNAs was strongly enriched in leaves and seedlings compared with their linear counterparts (**Figure 6**). In roots, and flowers, the ratios between the levels of circRNAs and corresponding linear RNAs were similar. The results of organs/



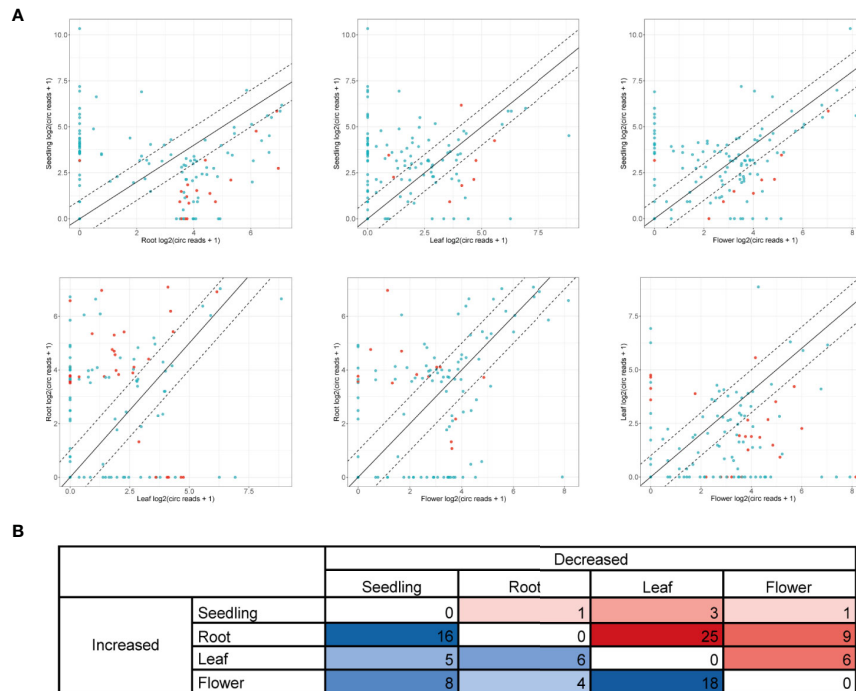


FIGURE 4 | (A) Pairwise comparison of circRNA accumulation in seedlings, roots, leaves, and flowers. Dashed line: two-fold cutoff; comparisons with p-value ≤ 0.05 marked in red. **(B)** The numbers of circRNAs with at least a two-fold increase or decrease in accumulation level between sample types (p-value ≤ 0.05).

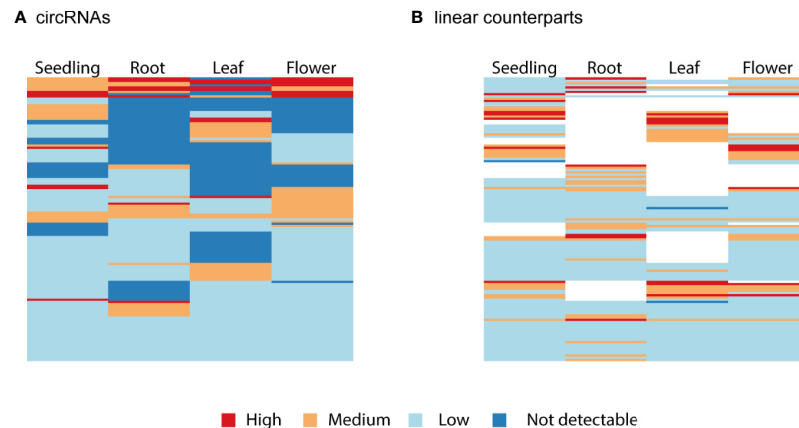
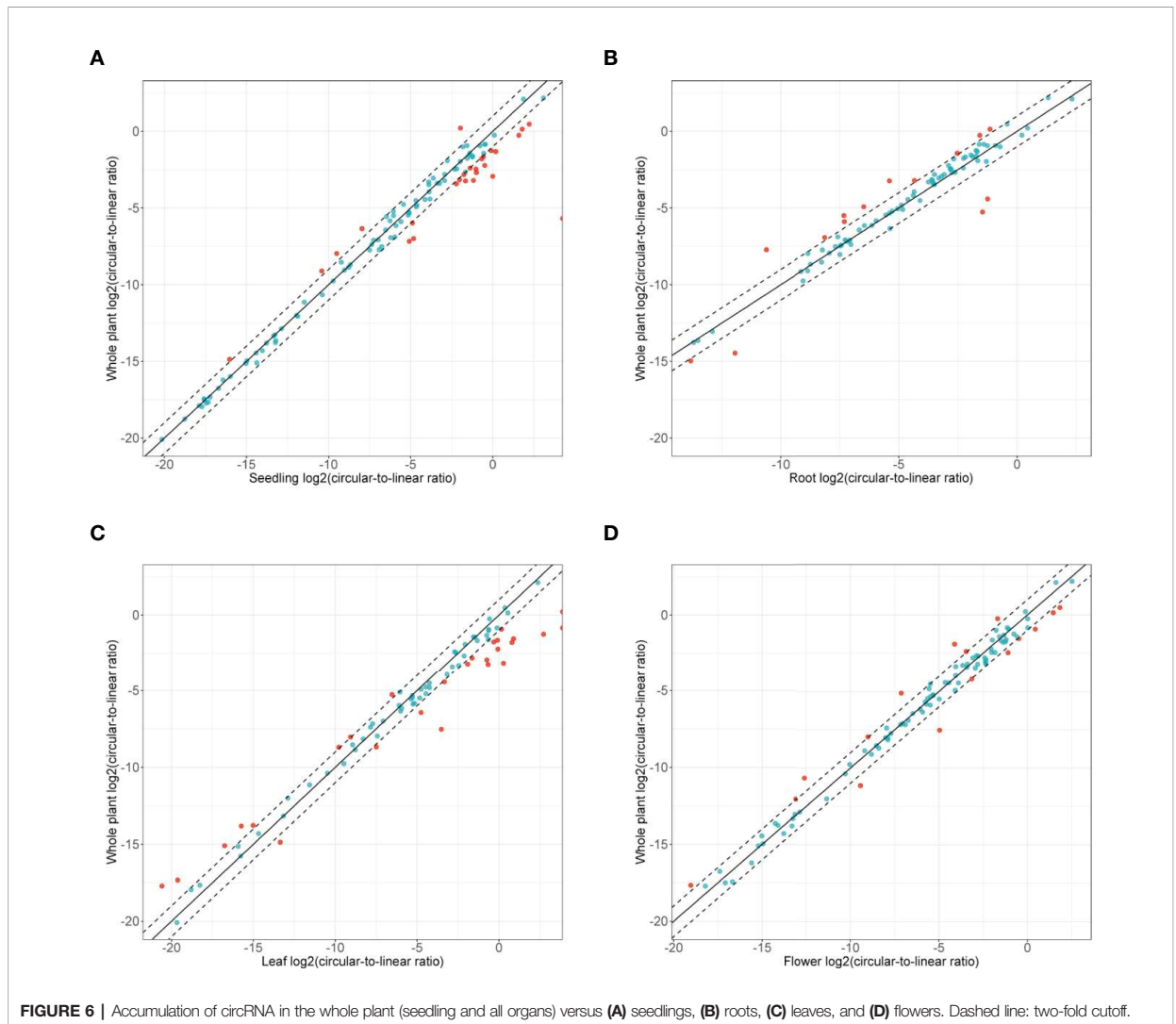


FIGURE 5 | Clustered heatmaps showing the accumulation pattern of the **(A)** most abundant circRNAs and **(B)** their linear counterparts in matching order. The Pearson correlation coefficient between circRNAs and linear transcripts was 0.46. When circRNA was not detected (dark blue), the linear isoform abundance was not assessed (white).

seedlings pairwise comparisons of circular to linear ratios are presented in **Additional file 2: Supplementary Figure 5**.

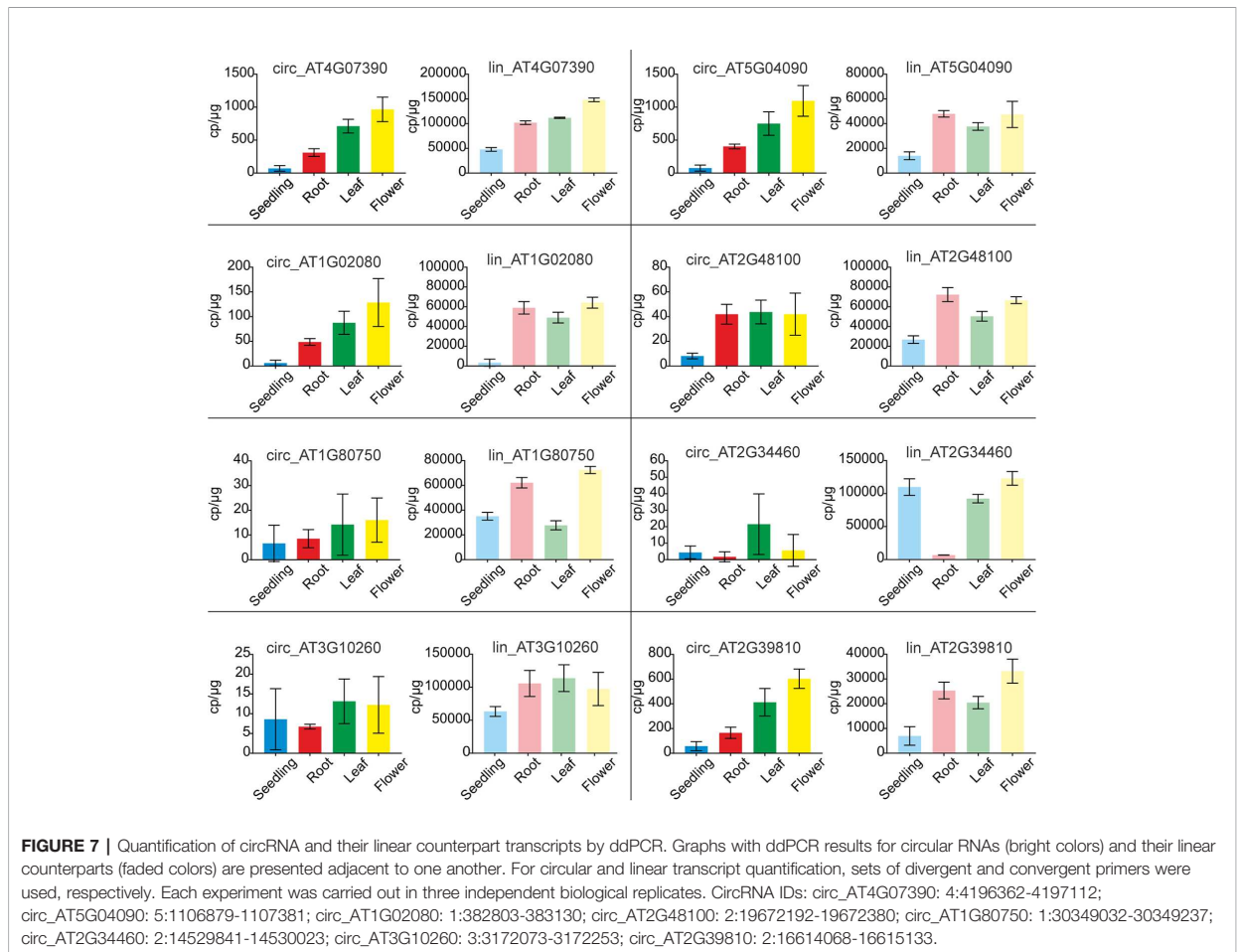
To verify the above observations, the levels of circRNAs and corresponding linear transcripts accumulation in seedlings, roots,

leaves, and flowers were precisely determined with ddPCR. To this end we selected 8 pairs of circular-linear RNAs. The most abundant circRNAs were circ_AT4G07390 and circ_AT5G04090, followed by circ_AT2G39810, circ_AT1G02080, circ_AT2G48100,



circ_AT2G34460, circ_AT1G80750, and circ_AT3G10260. Four of the tested circRNAs (circ_AT4G07390, circ_AT5G04090, circ_AT1G02080, and circ_AT2G39810) displayed a common accumulation profile, characterized by a continual increase in their amounts from seedlings through roots, leaves to flowers (Figure 7). In the case of circ_AT4G07390 and circ_AT2G39810, there was at least a 2-fold increase in the accumulation levels between seedlings and roots, as well as between roots and leaves, with a further increase in flowers. The other 4 analyzed circRNAs (circ_AT2G48100, circ_AT1G80750, circ_AT2G34460, and circ_AT3G10260) were rare, and in general, they did not reach 50 cp/μg. Therefore, the trends in their accumulation across plant organs and seedlings could not be reliably established. Lin_AT4G07390, the linear counterpart of one of the most abundant circRNA (circ_AT4G07390), had the highest expression

among the selected linear transcripts (Figure 7). Moreover, the expression level of lin_AT4G07390 increased from seedlings through roots, leaves, and flowers, which resembled the profile observed for circ_AT4G07390. The other two most abundant transcripts were, however, lin_AT3G10260 and lin_AT2G34460, the linear counterparts of circ_AT3G10260 and circ_AT2G34460, respectively. Notably, circ_AT3G10260 and circ_AT2G34460 were the rarest among the tested circRNAs. Similar discrepancies between the levels of circular and linear RNAs were detected for transcripts from AT5G04090 gene, whereas circ_AT5G04090 was one of the most abundant circRNAs, and its linear counterpart lin_AT5G04090 reached only medium expression values, comparable to lin_AT2G48100 and lin_AT1G80750 (the circular isoforms of which, circ_AT2G48100 and circ_AT1G80750, respectively, accumulated to very low levels). Taken together,



these data indicated that the abundance of circRNAs was not a simple consequence of the levels of parent gene expression (Pearson correlation coefficient: 0.14). Moreover, we identified 2 circRNAs, circ_AT5G04090 and circ_AT2G39810, that accumulated to considerable levels in all types of samples, and the profile of their accumulation was not correlated with the profile of linear transcript expression. These circRNAs were increased in leaves and flowers, and thus, we considered them organ-specific.

DISCUSSION

RNA-seq-based identification and quantification of circRNAs poses methodological challenges, which are plausible reasons for the substantial discrepancies in the numbers of individual circRNAs identified in *A. thaliana* in different studies (Wang et al., 2014; Ye et al., 2015). To address this issue, we tested how different experimental procedures and RNA-seq data analysis methods might influence circRNA profiling. First, we compared

two commonly used approaches for the preparation of sequencing libraries dedicated to circRNA identification. The first approach (R-) included only rRNA depletion, and the second one (R+) involved additional removal of linear transcripts by a 3' exonuclease, RNase R. RNase R treatment has been proposed to increase the efficiency of circRNA detection (Jeck et al., 2013) because it enables the enrichment of circRNAs, which otherwise would be missed in RNA-seq due to a large excess of linear transcripts. Our results were consistent with the above opinion. Although in the R+ libraries the average number of reads was approximately half of those in the R- libraries, the number of identified circRNA candidates was three times higher. Thus, the RNase R treatment clearly increased the resolution of the analyses. On the other hand, we found that some circRNAs, which were identified in four biological replicates in R- libraries and were subsequently validated by RT-PCR, were not identified in R+ libraries. This observation can be explained by the plausible nonenzymatic hydrolysis of some circRNAs upon incubation in reaction buffer, which renders them linear and exposes their 3' ends to RNase R digestion. Nonenzymatic

hydrolysis of RNA is a well-known phenomenon that occurs in the presence of multiple factors, including magnesium ions, polymeric organic compounds such as polyvinylpyrrolidone (PVP) and proteins (Bibillo et al., 1999; Bibillo et al., 2000). The hydrolysis leaves a 2',3'-cyclic phosphate at the 5' end of the cleavage side and a 5'-hydroxyl group at the 3' end (Kierzek, 1992). Because RNase R can accommodate 3'-phosphate-terminated substrates (Cheng and Deutscher, 2002), the products of nonenzymatic RNA hydrolysis can be digested by this enzyme. Thus, detailed biochemical studies of the factors that affect the stability of circRNAs upon incubation with RNase R are required to better understand to what extent reaction conditions impact the composition of circRNA pool. Taken together, RNase R treatment enriches circRNAs, however this procedure may also lead to the loss of some circRNAs. Consequently, we decided that the R- approach was more suitable for the identification and quantification of abundant circRNAs. In addition, regardless of the library type, we observed poor reproducibility of the circRNAs across biological replicates (ranging from 1.3 to 5.0%). The results suggested that the pool of circRNAs contained a substantial fraction of stochastically generated molecules, the physiological relevance of which is likely limited.

Our analyses revealed that the use of RNA-seq data for comparisons of different plant organs requires special attention for several reasons. Obviously, an indispensable prerequisite in such comparative studies is the normalization of RNA-seq data. In our study, the commonly used normalization by the library size was found to be inadequate because the obtained data poorly correlated with the ddPCR results (Pearson correlation coefficient: 0.67). The plausible reason for this result is the natural variation of the transcriptome composition of different organs. Consequently, the applied experimental procedures have different efficiencies, depending on the type of plant material. This phenomenon was evidenced by the varying effectiveness of rRNA depletion from particular organs. Moreover, the organ-specific differences may influence the RNA-seq ability to identify and quantify transcripts, in particular, the less abundant ones, such as circRNAs. To optimize the comparative analyses of circRNAs, we tested five other normalization methods. Four of these methods, namely, normalization by the (i) number of back-spliced reads, (ii) DESeq2 size factor (iii) library size without rRNA and chloroplast RNA reads and (iv) housekeeping gene *ACT2*, reached a Pearson correlation coefficient between the ddPCR and RNA-seq results of 0.76 to 0.77. In the case of single-organ quantitative analyses of circRNA, the correlation was even higher, reaching 0.92. This observation has two implications. First, it supports the opinion that the transcriptome composition has a considerable impact on RNA-seq-based profiling of rare RNA species. Second, this result demonstrates the high reliability of RNA-seq-based quantitative characterization of circRNAs within a particular organ, thus corroborating the results from some previous studies that were not validated by qPCR (Ye et al., 2015; Dou et al., 2017). Moreover, this finding opens a perspective for large-scale single-organ quantitative studies of circRNAs

under various conditions and comparisons of their accumulation levels in response to different stimuli.

Our analysis of circRNAs across *A. thaliana* seedlings and organs was based on 127 circRNAs identified by CIRI2 and find_circ programs and supported by at least 10 normalized RNA-seq reads, which corresponded to 50 copies per microgram of total RNA. Based on the provided data, we believe that this set well represents reproducibly generated circRNAs that are most likely to be physiologically relevant. In general, circRNAs were accumulated in an organ-specific manner; however, approximately 41% were ubiquitous. GO analysis revealed that genes giving rise to circRNAs were significantly enriched in those involved in photosynthesis in seedlings, leaves and flowers. However, roots had the highest number of unique circRNAs and most circRNAs of significantly increased accumulation (p -value ≤ 0.05) in comparison to other organs/seedlings. The smallest number of circRNAs with increased accumulation was found for the seedlings. Moreover, our results clearly demonstrated that there was no direct link between the accumulation of circRNAs and their linear counterparts. While the levels of some circRNAs increased with the increase in their linear counterpart transcripts (circ_AT4G07390 and lin_AT4G07390), there was an overall lack of correlation between the accumulation of these two transcriptome fractions (Pearson correlation coefficient: 0.14). This finding indicated that there have to exist specific mechanisms regulating the formation of reproducibly generated circRNAs. What is more, some circRNAs (for example, circ_AT5G04090 and circ_AT2G39810) were selectively increased in particular organs, while the levels of their linear counterparts remained relatively stable throughout the plant. It is tempting to speculate that their biogenesis involves an organ-specific molecular switch that changes the regular splicing pattern. These circRNAs also emerge as attractive candidates for functional studies.

Recently, chloroplasts were described as circRNA production hotspots, and circRNAs were proposed to be involved in photosynthesis (Sun et al., 2016). Our data revealed that back-spliced reads supporting circRNAs derived from the chloroplast genome were indeed prevalent in leaves. These findings raise intriguing questions about how such substantially different processes as nuclear and organellar splicing, could both lead to the formation of circRNAs. However, this issue is complicated by the observation that some organellar mRNAs are products of trans-splicing, in which the exons of two primary transcripts are joined (de Longevialle et al., 2007; Aryamanesh et al., 2017). Some reads derived from such mRNAs can be misclassified as back-spliced reads. In addition, chloroplast group II introns can be excised as full-length circles instead of lariats (Lasda and Parker, 2014). In the above context, chloroplast circRNAs identified by RNA-seq require extensive validation.

Here, we integrated RNA-seq and ddPCR to study *A. thaliana* circRNAs. By using ddPCR results as feedback to improve RNA-seq normalization, we developed a reliable strategy of circRNA comparative analyses. This strategy allowed us to comprehensively characterize circRNAs across plant organs and in seedlings. Our results stimulate further research of organ-specific circRNAs and of

the mechanisms that guide a switch between the formation of circular and their linear counterpart transcripts.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA525820>.

AUTHOR CONTRIBUTIONS

MF conceived the overall idea of the study. AP, PJ, and MF conceived and designed the experiments. KN performed all bioinformatics analyses. AP, PJ and MF analyzed and discussed the results, MS cultivated plants, isolated RNA and performed PCR/ddPCR. MS and PJ cloned and sequenced PCR products. MS and JP prepared RNA samples for sequencing. JP and LH generated the RNA-seq libraries. AP, KN, MS, and PJ drafted the manuscript. MF

REFERENCES

- Aryamanesh, N., Ruwe, H., Sanglard, L. V., Eshraghi, L., Bussell, J. D., Howell, K. A., et al. (2017). The Pentatricopeptide Repeat Protein EMB2654 Is Essential for Trans-Splicing of a Chloroplast Small Ribosomal Subunit Transcript. *Plant Physiol.* 173 (2), 1164–1176. doi: 10.1104/pp.16.01840
- Bibillo, A., Figlerowicz, M., and Kierzek, R. (1999). The non-enzymatic hydrolysis of oligoribonucleotides VI. The role of biogenic polyamines. *Nucleic Acids Res.* 27 (19), 3931–3937. doi: 10.1093/nar/27.19.3931
- Bibillo, A., Figlerowicz, M., Ziomek, K., and Kierzek, R. (2000). The nonenzymatic hydrolysis of oligoribonucleotides. VII. Structural elements affecting hydrolysis. *Nucleosides Nucleotides Nucleic Acids* 19 (5-6), 977–994. doi: 10.1080/15257770008033037
- Boyes, D. C., Zayed, A. M., Ascenzi, R., McCaskill, A. J., Hoffman, N. E., Davis, K. R., et al. (2001). Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *Plant Cell* 13 (7), 1499–1510. doi: 10.1105/TPC.010011
- Chen, G., Cui, J., Wang, L., Zhu, Y., Lu, Z., and Jin, B. (2017). Genome-Wide Identification of Circular RNAs in Arabidopsis thaliana. *Front. Plant Sci.* 8, 1678. doi: 10.3389/fpls.2017.01678
- Chen, L., Zhang, P., Fan, Y., Lu, Q., Li, Q., Yan, J., et al. (2018). Circular RNAs mediated by transposons are associated with transcriptomic and phenotypic variation in maize. *New Phytol.* 217 (3), 1292–1306. doi: 10.1111/nph.14901
- Cheng, Z. F., and Deutscher, M. P. (2002). Purification and characterization of the Escherichia coli exoribonuclease RNase R. Comparison with RNase II. *J. Biol. Chem.* 277 (24), 21624–21629. doi: 10.1074/jbc.M202942200
- Cheng, J., Zhang, Y., Li, Z., Wang, T., Zhang, X., and Zheng, B. (2018). A lariat-derived circular RNA is required for plant development in Arabidopsis. *Sci. China Life Sci.* 61 (2), 204–213. doi: 10.1007/s11427-017-9182-3
- Chu, Q., Zhang, X., Zhu, X., Liu, C., Mao, L., Ye, C., et al. (2017). PlantcircBase: A Database for Plant Circular RNAs. *Mol. Plant* 10 (8), 1126–1128. doi: 10.1016/j.molp.2017.03.003
- Darbani, B., Noeparvar, S., and Borg, S. (2016). Identification of Circular RNAs from the Parental Genes Involved in Multiple Aspects of Cellular Metabolism in Barley. *Front. Plant Sci.* 7, 776. doi: 10.3389/fpls.2016.00776
- de Longevialle, A. F., Meyer, E. H., Andres, C., Taylor, N. L., Lurin, C., Millar, A. H., et al. (2007). The pentatricopeptide repeat gene OTP43 is required for trans-splicing of the mitochondrial nad1 Intron 1 in Arabidopsis thaliana. *Plant Cell* 19 (10), 3256–3265. doi: 10.1105/tpc.107.054841
- Dheda, K., Huggett, J. F., Bustin, S. A., Johnson, M. A., Rook, G., and Zumla, A. (2004). Validation of housekeeping genes for normalizing RNA expression in

was responsible for the final version of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financed by the National Science Centre grant number UMO-2014/15/D/NZ2/02305 to AP. Partial financial support was also provided by the Polish Ministry of Science and Higher Education under the KNOW program. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2020.576581/full#supplementary-material>

- real-time PCR. *Biotechniques* 37 (1), 112–4, 116, 118–9. doi: 10.2144/04371RR03
- Dou, Y., Li, S., Yang, W., Liu, K., Du, Q., Ren, G., et al. (2017). Genome-wide Discovery of Circular RNAs in the Leaf and Seedling Tissues of Arabidopsis Thaliana. *Curr. Genomics* 18 (4), 360–365. doi: 10.2174/1389202918666170307161124
- Gao, Y., Wang, J., and Zhao, F. (2015). CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol.* 16, 4. doi: 10.1186/s13059-014-0571-3
- Gao, Y., Zhang, J., and Zhao, F. (2017). Circular RNA identification based on multiple seed matching. *Brief. Bioinform.* 19 (5), 803–810. doi: 10.1093/bib/bbx014
- Hansen, T. B. (2018). Improved circRNA Identification by Combining Prediction Algorithms. *Front. Cell Dev. Biol.* 6, 20. doi: 10.3389/fcell.2018.00020
- Jeck, W. R., Sorrentino, J. A., Wang, K., Slevin, M. K., Burd, C. E., Liu, J., et al. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 19 (2), 141–157. doi: 10.1261/rna.035667.112
- Kierzek, R. (1992). Nonenzymatic hydrolysis of oligoribonucleotides. *Nucleic Acids Res.* 20 (19), 5079–5084. doi: 10.1093/nar/20.19.5079
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357–360. doi: 10.1038/nmeth.3317
- Kopylova, E., Noe, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28 (24), 3211–3217. doi: 10.1093/bioinformatics/bts611
- Lasda, E., and Parker, R. (2014). Circular RNAs: diversity of form and function. *RNA* 20 (12), 1829–1842. doi: 10.1261/rna.047126.114
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi: 10.1093/bioinformatics/btp324
- Liu, T., Zhang, L., Chen, G., and Shi, T. (2017). Identifying and Characterizing the Circular RNAs during the Lifespan of Arabidopsis Leaves. *Front. Plant Sci.* 8, 1278. doi: 10.3389/fpls.2017.01278
- Lu, T., Cui, L., Zhou, Y., Zhu, C., Fan, D., Gong, H., et al. (2015). Transcriptome-wide investigation of circular RNAs in rice. *RNA* 21 (12), 2076–2087. doi: 10.1261/rna.052282.115
- Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495 (7441), 333–338. doi: 10.1038/nature11928
- Pan, T., Sun, X., Liu, Y., Li H., Deng, G., Lin, H., et al. (2018). Heat stress alters genome-wide profiles of circular RNAs in Arabidopsis. *Plant Mol. Biol.* 96 (3), 217–229. doi: 10.1007/s11103-017-0684-7

- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33 (3), 290–295. doi: 10.1038/nbt.3122
- Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11 (9), 1650–1667. doi: 10.1038/nprot.2016.095
- Qu, S., Liu, Z., Yang, X., Zhou, J., Yu, H., Zhang R., et al. (2018). The emerging functions and roles of circular RNAs in cancer. *Cancer Lett.* 414, 301–309. doi: 10.1016/j.canlet.2017.11.022
- Radonic, A., Thulke, S., Mackay, I. M., Landt, O., Siegert, W., and Nitsche, A. (2004). Guideline to reference gene selection for quantitative real-time PCR. *Biochem. Biophys. Res. Commun.* 313 (4), 856–862. doi: 10.1016/j.bbrc.2003.11.177
- Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L., and Brown, P. O. (2013). Cell-type specific features of circular RNA expression. *PLoS Genet.* 9 (9), e1003777. doi: 10.1371/annotation/tf782282b-eefa-4c8d-985c-b1484e845855
- Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* 9, 88. doi: 10.1186/s13104-016-1900-2
- Sun, X., Wang, L., Ding, J., Wang, Y., Wang, J., Zhang, X., et al. (2016). Integrative analysis of *Arabidopsis thaliana* transcriptomics reveals intuitive splicing mechanism for circular RNA. *FEBS Lett.* 590 (20), 3510–3516. doi: 10.1002/1873-3468.12440
- Szabo, L., and Salzman, J. (2016). Detecting circular RNAs: bioinformatic and experimental challenges. *Nat. Rev. Genet.* 17 (11), 679–692. doi: 10.1038/nrg.2016.114
- Tan, J., Zhou, Z., Niu, Y., Sun, X., and Deng, Z. (2017). Identification and Functional Characterization of Tomato CircRNAs Derived from Genes Involved in Fruit Pigment Accumulation. *Sci. Rep.* 7 (1), 8594. doi: 10.1038/s41598-017-08806-0
- Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45 (W1), W122–W129. doi: 10.1093/nar/gkx382
- Wang, P. L., Bao, Y., Yee, M. C., Barrett, S. P., Hogan, G. J., Olsen, M. N., et al. (2014). Circular RNA is expressed across the eukaryotic tree of life. *PLoS One* 9 (6), e90859. doi: 10.1371/journal.pone.0090859
- Wang, Y., Yang, M., Wei, S., Qin, F., Zhao, H., and Suo B. (2016). Identification of Circular RNAs and Their Targets in Leaves of Triticum aestivum L. under Dehydration Stress. *Front. Plant Sci.* 7, 2024. doi: 10.3389/fpls.2016.02024
- Ye, C. Y., Chen, L., Liu C., Zhu, Q. H., and Fan, L. (2015). Widespread noncoding circular RNAs in plants. *New Phytol.* 208 (1), 88–95. doi: 10.1111/nph.13585
- Ye, J., Wang, L., Li, S., Zhang, Q., Zhang, Q., Tang, W., et al. (2017). AtCircDB: a tissue-specific database for *Arabidopsis* circular RNAs. *Brief. Bioinform.* 20 (1), 58–65. doi: 10.1093/bib/bbx089
- Zhao, T., Wang, L., Li, S., Xu, M., Guan, X., and Zhou, B. (2017). Characterization of conserved circular RNA in polyploid *Gossypium* species and their ancestors. *FEBS Lett.* 591 (21), 3660–3669. doi: 10.1002/1873-3468.12868
- Zuo, J., Wang, Q., Zhu, B., Luo, Y., and Gao, L. (2016). Deciphering the roles of circRNAs on chilling injury in tomato. *Biochem. Biophys. Res. Commun.* 479 (2), 132–138. doi: 10.1016/j.bbrc.2016.07.032

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Philips, Nowis, Stelmaszczuk, Jackowiak, Podkowiński, Handschuh and Figlerowicz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

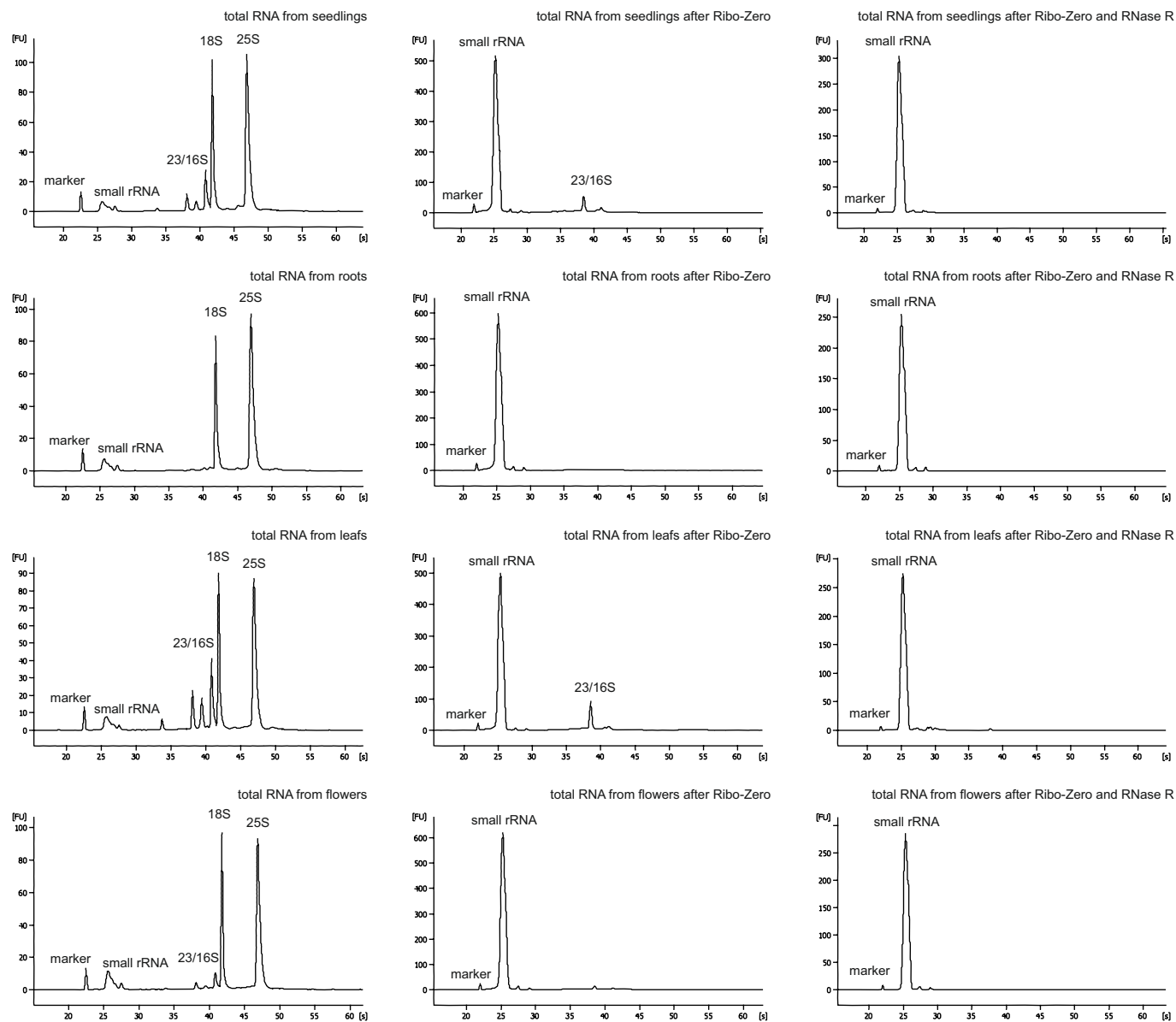
Materiały Suplementarne

Philips A, **Nowis K**, Stelmaszczuk M, Jackowiak P, Podkowiński J, Handschuh L,
Figlerowicz M;

Expression Landscape of circRNAs in Arabidopsis thaliana Seedlings and Adult
Tissues

Frontiers in Plant Science (2020), <https://doi.org/10.3389/fpls.2020.576581>

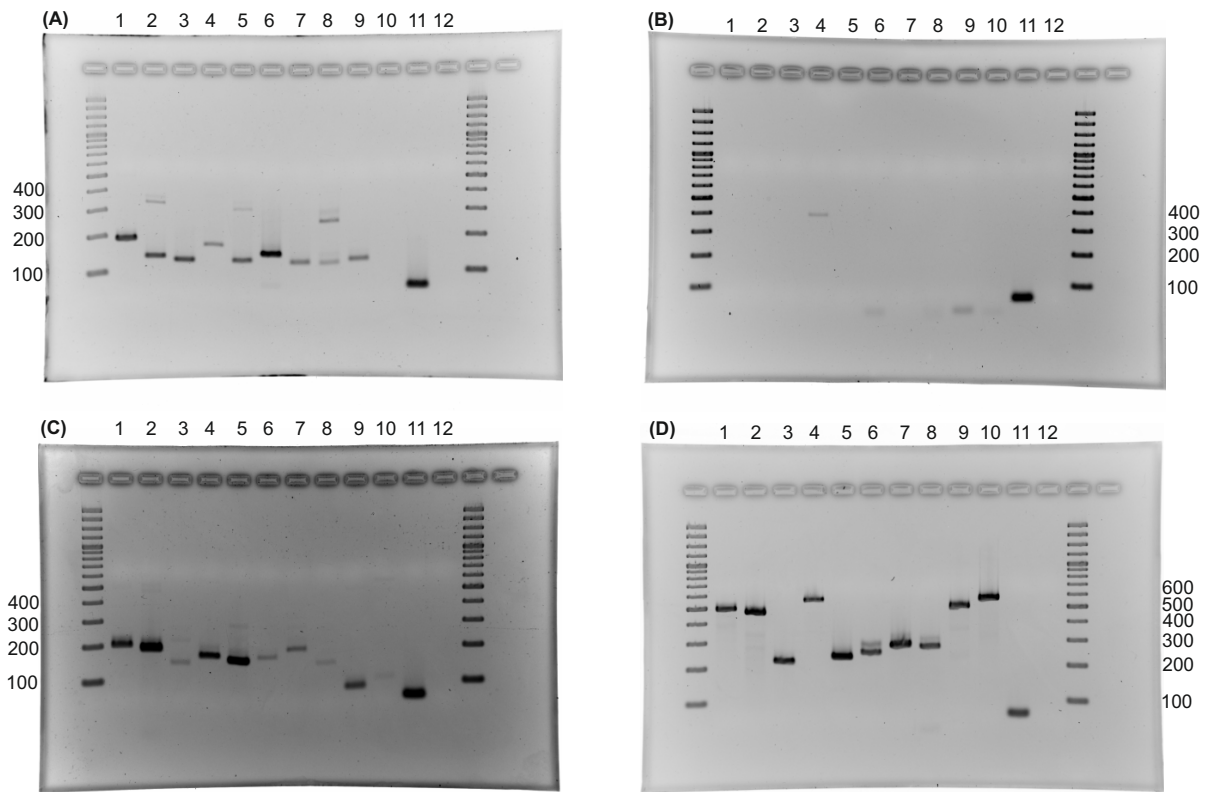
Pięcioletni IF: 7,3, MEiN 100



Supplementary Figure 1.

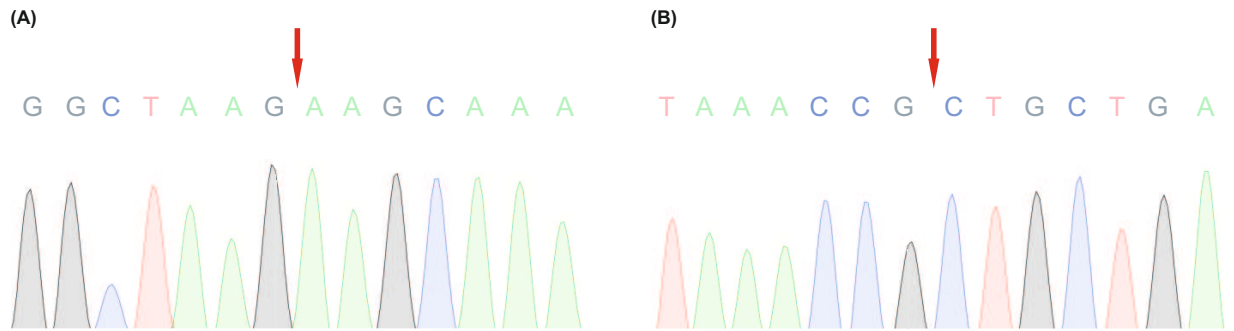
The efficiency of Ribo-Zero and RNase R treatment of total RNA from seedlings and plant organs.

Representative electropherograms (Bioanalyzer, Agilent) of total RNAs (left side), total RNAs after Ribo-Zero treatment (middle), and total RNAs after Ribo-Zero and RNase R treatment (right side) from seedlings, roots, leaves, and flowers. One hundred nanograms of DNase-treated total RNA of each sample was analyzed on the Bioanalyzer, Agilent with RNA Nano Chip in accordance with the manufacturer's instructions. After Ribo-Zero treatment, 3 ng of each sample was analyzed on the Bioanalyzer, Agilent with RNA Pico Chip in accordance with the manufacturer's instructions. After Ribo-Zero and RNase R treatment, 3 ng of each sample was analyzed on the Bioanalyzer, Agilent with RNA Pico Chip in accordance with the manufacturer's instructions. The x-axis represents the resolving time in seconds (s), and the y-axis represents fluorescence (fu). In addition to the cytoplasmic 25S and 18S rRNA peaks, other peaks corresponding to 23S and 16S rRNA from chloroplasts and small rRNAs are present.



Supplementary Figure 2.
 Uncropped gel images used in Figure 1 in the main text.

- (A) Uncropped gel image used in panel B, Figure 1.
- (B) Uncropped gel image used in panel C, Figure 1.
- (C) Uncropped gel image used in panel D, Figure 1.
- (D) Uncropped gel image used in panel E, Figure 1.



(C)

```
CCGGCAAGAAGAAGAAGAAAGGCGTAAAAGCCGCTTTTGGTTTGCCTTGGTCAGCAAGGTACAAGGTAGCGTT
AGGTATAGCAGATGCCATAGCTTATTTGCATAATGGAACAGAGCAATGTGTTGTTTCACAGAGATATCAAACCCTC
CAATATTCTTCTTTCTCAAAGAAAATACCAAAGGTAAGATTCGCGAAACGAATGATGTATTCGATAATTATCGGT
AAGCAAACCAATGATTTGTTTTGACCTGTTCTTTTCAGTTGTGTGATTTGGGTTAGCTACTTGGACCGCTGCA
CCTTCGGTTCCTTTCCTCTGCAAGACCGTGAAAGGCACATTCGGGTATTGCTCGAAACTTGCCTTGAGTTGTTT
CTTCTTCTTCTTGCCGGA
```

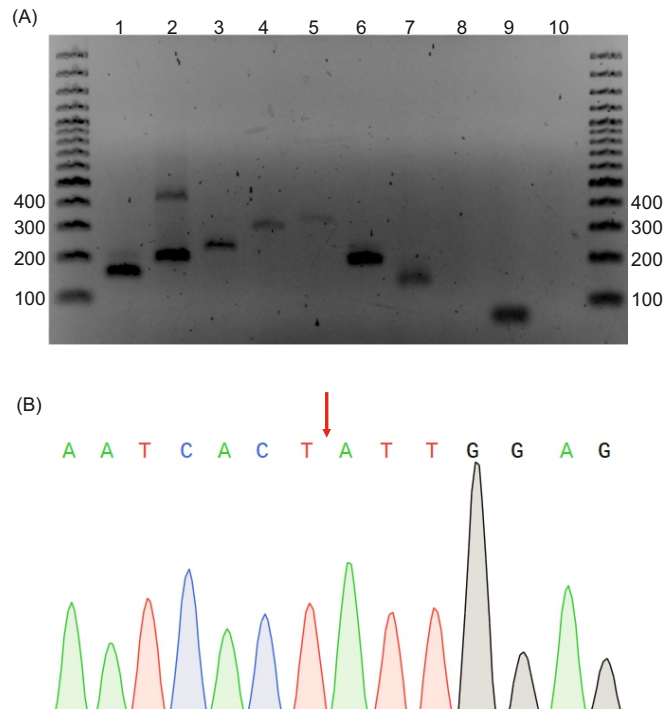
Supplementary Figure 3.

Results of Sanger sequencing of PCR products.

(A) Sequencing chromatogram of the circRNA 4 (1:30349032-30349237) amplification product (Figure 1, panel B, lane 4). The red arrow indicates the back-splice site.

(B) Sequencing chromatogram of the circRNA 7 (3:3172073-3172253) amplification product (Figure 1, panel B, lane 7). The red arrow indicates the back-splice site.

(C) The sequence of the mispriming amplification product (Figure 1, panel C, lane 4). It was identified as a fragment of the AT5G37790 locus.

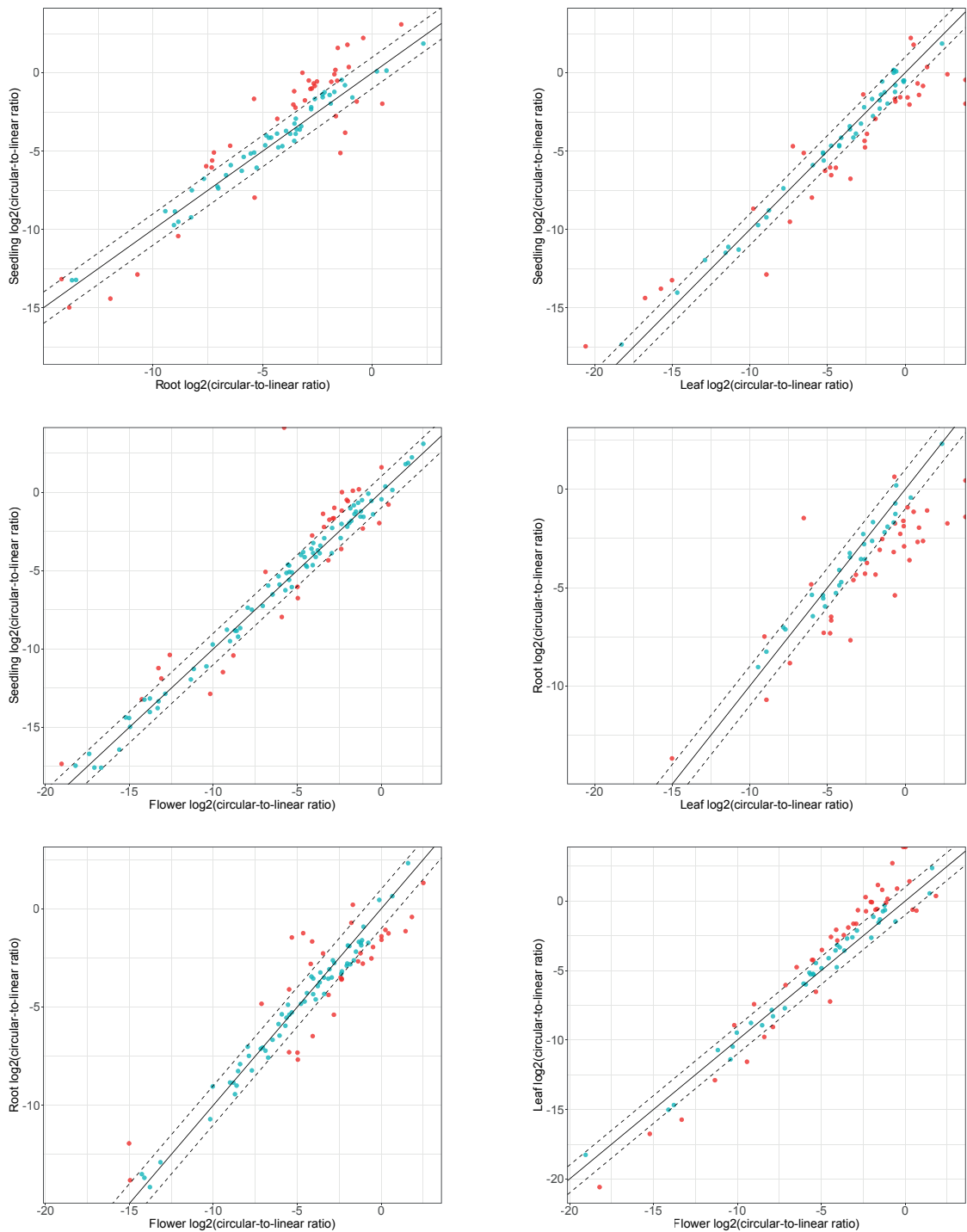


Supplementary Figure 4.

(A) Detection of circRNAs not identified in the R+ libraries. RT-PCR with cDNA from the leaf as template was performed as described in the main text. For PCRs 1-8, divergent (circular specific) primers were used. Lanes 1 to 6 show PCR products corresponding to circRNA candidates that were not identified in the R+ libraries. Lane 7 shows a circRNA derived from the chloroplast chromosome that was found in both libraries. Lane 8 shows a circRNA that was not identified in the R- library in leaves but was detected in seedlings. Lanes 9 and 10 represent positive (leaf cDNA template) and negative (no template) controls, respectively, with convergent primers for ACT2.

CircRNA IDs: 1) 1:20310304-20310540; 2) 4:2719063-2719462; 3) 1:7085326-7088687; 4) 1:26715933-26716400; 5) 1:20066441-20074643; 6) 1:22359475-22363105; 7) Pt:419-556; 8) 3:4922906-4925556.

(B) Sequencing chromatogram of the circRNA 6 (1:22359475-22363105) amplification product (panel A, lane 6). The red arrow indicates the back-splice site.



Supplementary Figure 5.

The organs/seedling pairwise comparison of circular to linear ratios. Dashed line: 2-fold cut-off.

2. Philips A, **Nowis K**, Stelmaszczuk M, Podkowiński J, Handschuh L, Jackowiak P, Figlerowicz M;



Arabidopsis thaliana cbp80, c2h2, and flk Knockout Mutants Accumulate
Increased Amounts of Circular RNAs

Cells (2020), <https://doi.org/10.3390/cells9091937>

Pięcioletni IF 7,7, MEiN 140

Article

Arabidopsis thaliana *cbp80*, *c2h2*, and *flk* Knockout Mutants Accumulate Increased Amounts of Circular RNAs

Anna Philips¹, Katarzyna Nowis¹, Michal Stelmaszczuk¹, Jan Podkowiński¹,
Luiza Handschuh¹ , Paulina Jackowiak^{1,*} and Marek Figlerowicz^{1,2,*} 

¹ Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61-704 Poznan, Poland; aphilips@ibch.poznan.pl (A.P.); kkozłowska@ibch.poznan.pl (K.N.); michal.stelmaszczuk@ibch.poznan.pl (M.S.); jpodkow@man.poznan.pl (J.P.); luizahan@ibch.poznan.pl (L.H.)

² Institute of Computing Science, Poznan University of Technology, 60-965 Poznan, Poland

* Correspondence: paulinaj@ibch.poznan.pl (P.J.); marekf@ibch.poznan.pl (M.F.)

Received: 17 July 2020; Accepted: 18 August 2020; Published: 21 August 2020



Abstract: Circular RNAs (circRNAs) are the products of the non-canonical splicing of pre-mRNAs. In contrast to humans and animals, our knowledge of the biogenesis and function of circRNAs in plants is very scarce. To identify proteins involved in plant circRNA generation, we characterized the transcriptomes of 18 *Arabidopsis thaliana* knockout mutants for genes related to splicing. The vast majority (>90%) of circRNAs were formed in more than one variant; only a small fraction of circRNAs was mutant-specific. Five times more circRNA types were identified in *cbp80* and three times more in *c2h2* mutants than in the wild-type. We also discovered that in *cbp80*, *c2h2* and *flk* mutants, the accumulation of circRNAs was significantly increased. The increased accumulation of circular transcripts was not accompanied by corresponding changes in the accumulation of linear transcripts. Our results indicate that one of the roles of CBP80, C2H2 and FLK in splicing is to ensure the proper order of the exons. In the absence of one of the above-mentioned factors, the process might be altered, leading to the production of circular transcripts. This suggests that the transition toward circRNA production can be triggered by factors sequestering these proteins. Consequently, the expression of linear transcripts might be regulated through circRNA production.

Keywords: circRNA; *Arabidopsis thaliana*; splicing; *cbp80*; *c2h2*; *flk*; RNA-seq

1. Introduction

Circular RNAs (circRNAs) are a class of non-coding alternatively spliced RNA transcripts. It has been shown that circRNAs are present across the eukaryotic tree of life [1]. Moreover, many of them are evolutionarily conserved and highly abundant [2,3]. These facts, together with the newest findings on some circRNA functions [4–6], strongly suggest important the roles of circRNAs in crucial cellular processes, especially in the regulation of gene expression [7]. In humans and animals, circRNAs may act as protein-binding molecules [7–9] or miRNA sponges [6,9,10]. For example, CDR1as/ciRS-7 and *Sry* circRNAs have many miRNA-binding sites, via which they sequester miRNA from its target sites in mRNA. It is also becoming clear that many circRNAs are involved in the pathogenesis of numerous diseases, e.g., cancer [11,12] or Alzheimer's disease [13]. In addition, several mechanisms of circRNA formation have been postulated. It has been shown that circRNA generation is to some degree dependent on the length and sequence of introns [2,7,14]; however, there are many exceptions to this general rule [1]. It is also known that circRNAs are generated in a cell-specific manner, and their accumulation levels are not correlated with those of corresponding mRNAs [15]. The facts listed above permit us to hypothesize that in addition to the protein/RNA factors involved in canonical

splicing, there are other unknown factors that promote circRNA generation. Moreover, as circRNA and mRNA are produced from the same exons by the same machinery (spliceosome), it is highly probable that circRNAs may function in the regulation of gene expression by competing with linear mRNA generation [7]. All the above-mentioned observations come from the study of circRNA in animals and humans. Unfortunately, so far, little is known about the biogenesis and roles of circRNAs in plants. To date, only one plant circRNA has been clearly assigned functions; circRNA derived from exon 6 of the *SEPALLATA3* (*SEP3*) gene was shown to regulate flower development in *Arabidopsis thaliana* [16].

To identify genetic factors that influence circRNA production, we used one of the best-studied model dicot plants, *A. thaliana*. This plant has thousands of easily available mutants, including well-characterized single-gene knockouts. Considering the putative mechanisms of circRNA formation, we selected wild-type *A. thaliana* (Columbia ecotype, abbreviated Col-0) and its 18 variants, each harboring a knockout mutation in a gene that encodes a protein involved in splicing [17]. The compositions of the transcriptomes of all these plants were determined via RNA-seq. Consequently, genes potentially affecting circRNA production were identified.

2. Materials and Methods

2.1. Plant Material and RNA Extraction

Seeds of Col-0 and T-DNA insertion knockout mutants of *A. thaliana* plants were obtained from Arabidopsis Biological Resource Center (ABRC). All the ABRC IDs (deposition numbers) with the names of knocked-out genes in mutant plants used in the study are listed in Table S1. Formal identification of the plant material used in the study was performed by the J. Ecker or Bernd Weisshaar laboratory and Syngenta (plant material donors) using Illumina sequencing or PCR.

The seeds of the *A. thaliana* plants were sterilized by placing them in filter tubes and washing with 70% ethanol, distilled water, 0.01% Amistar 250 SC (Syngenta, Warsaw, Poland), and again with distilled water using a vacuum pump. Next, the seeds were stratified for 4 days at 4 °C in 0.1% agarose solution. Plants were grown using Arasystem (Betatech, Ghent, Belgium) and Jiffy-7 peat pellets in a growth chamber with 16 h of light at 23 °C and 8 h of dark at 18 °C. Whole leaves at stage 3.90 (rosette growth complete) [18] were collected, frozen in liquid nitrogen, and stored at −80 °C until genotyping and RNA isolation.

Prior to obtaining RNA, mutant plants were genotyped using a strategy proposed by the Salk Institute Genomic Analysis Laboratory for *A. thaliana* T-DNA insertion mutants. Briefly, the T-DNA Primer Design Tool was used to obtain primer sequences for each mutant. All the primers (Genomed, Warsaw, Poland) used in genotyping reactions are listed in Table S1. Next, with a forward border primer (FBP) targeting the T-DNA insertion and forward (FP) and reverse (RP) primers that flanked the T-DNA insertion site, PCRs with genomic DNA (gDNA) from the leaves of mutant and Col-0 plants (as a control) were performed.

The PCR mixtures for the genotyping reaction (25 µL of the total volume) contained 10 ng of gDNA, 0.625 U of Taq DNA polymerase (5 U/µL), 2.5 mM MgCl₂, 0.2 mM dNTPs, 1× Taq Buffer with (NH₄)₂SO₄ (Thermo Fisher Scientific, Waltham, MA, USA), and each primer at 0.5 µM (Genomed, Warsaw, Poland). PCR was performed in a T-100 thermal cycler (Bio-Rad, Hercules, CA, USA) using the following program: preheating at 95 °C for 3 min, followed by 25 cycles of 95 °C for 30 s, 62 °C for 30 s and 72 °C for 1 min and 15 s. Finally, the reaction mixtures were subjected to elongation at 72 °C for 5 min. The QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) was used for the purification of PCR products, which were then resolved on a 2% agarose gel for 70 min at 120 V in a Wide Mini-Sub Cell GT System (Bio-Rad, Hercules, CA, USA), and subsequently stained with a 0.5 µg/mL EtBr solution for 20 min on a rocker. A ChemiDoc XRS+ System (Bio-Rad, Hercules, CA, USA) was used to obtain gel images, which were analyzed with Image Lab software (Bio-Rad, Hercules, CA, USA).

Total RNA was isolated from 100 mg of powdered leaf samples from 4 biological replicates for each mutant and wild-type plant using the mirVana miRNA Isolation Kit (Thermo

Fisher Scientific, Waltham, MA, USA). For each replicate, 10 plants were used to obtain powdered samples. Subsequently, 10 µg of total RNA was treated with 2 U of Turbo DNase (Thermo Fisher Scientific, Waltham, MA, USA) and purified with the QIAquick Nucleotide Removal Kit (Qiagen, Hilden, Germany). Capillary electrophoresis (2100 Bioanalyzer, Agilent, Santa Clara, CA, USA) with the Plant RNA Nano Assay was used to assess RNA quality and integrity. For further analyses, we used samples with an RNA integrity number (RIN) >7.5.

2.2. Library Preparation and Sequencing

For library preparation, we used 2.5 µg of total RNA (isolated as described above) treated with the Ribo-Zero rRNA Removal Kit (Plant Leaf) (Illumina, San Diego, CA, USA) following the manufacturer's recommendation. The level of rRNA depletion was determined by capillary electrophoresis (2100 Bioanalyzer, Agilent, Santa Clara, CA, USA) with the Plant RNA Pico Assay.

One hundred-nanogram aliquots of RNA samples were prepared in 4 biological replicates. Next, libraries were obtained separately using Kapa strand RNA-Seq library synthesis with NEB adapter modification. The libraries were subjected to qualitative analysis using capillary electrophoresis (2100 Bioanalyzer, Agilent, Santa Clara, CA, USA) with the High Sensitivity DNA Assay and quantitative analysis with a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA) prior to sequencing. The samples were sequenced with the Genome Analyzer IIx (Illumina) and 108-bp paired-end protocol.

2.3. CircRNA Identification and Quantification

Remnant adapter sequences were removed, low-quality bases at read ends were trimmed out (minimum quality score of 30; 99.9% base call accuracy) and reads shorter than 20 nucleotides were removed with AdapterRemoval (version 1.5.4; [19]).

CircRNAs were identified using a protocol proposed [20]. Briefly, filtered reads were mapped with BWA mem (version 0.7.10; [21]) to the reference genome (TAIR10), and based on back-spliced read identification, CIRI2 generated a list of circRNA candidates supported by at least 2 back-spliced reads. Additionally, the results obtained with CIRI2 were validated with find_circ [9].

Subsequently, the back-spliced reads supporting circRNAs were normalized to the library size of a sample reduced by reads mapped to chloroplasts and rRNA. Reads mapped to rRNA were determined using SortMeRNA (version 2.1; [22]). Plots were generated using the R package (version 3.6.1) with ggplot2 and wiggleplotr libraries.

2.4. Transcript-Level Expression and Alternative Splicing Analysis

Filtered reads were mapped to the reference genome (TAIR10) using HISAT2 (version 2.1.0; [23–25]). Transcript assembly and quantification was assessed with StringTie (version 1.3.3b [23]). Differential expression analysis of the quantified transcripts was made using the DESeq2 library (version 1.26.0; [26]). Plots were generated using the R package (version 3.6.1) with ggplot2 and wiggleplotr libraries.

3. Results

3.1. Selection of the *A. thaliana* Knock-Out Mutants

In this study, we analyzed circRNA and mRNA accumulation in wild-type *A. thaliana* and its selected mutants that had changes in the genes involved in splicing. We chose these mutants as the primary objects of our analyses because the newest findings indicate that transcript circularization is regulated by general splicing factors, cis-elements and cognate factors, but with unknown regulatory rules that are distinct from those of canonical splicing [1]. According to the Arabidopsis Splicing Related Genes (ASRG) database, 395 genes encode splicing-associated proteins in *A. thaliana* [17]. Preliminary *A. thaliana* variant selection led us to a list of 86 mutants with changes in genes related to splicing, of which 23 had a precisely described phenotype (data from the ABRC [27]). Furthermore, we narrowed

the set of mutants to those with a Col-0 background so as to obtain a common reference for all of the analyzed plants. We also excluded heterozygotes and variants with lethal mutations. As a result, the studied plants included wild-type *A. thaliana* and its 18 mutants listed in Table S1. All plants were cultivated and genotyped as described in the Materials and Methods.

3.2. CircRNA Identification in Wild-Type and Mutant Plants

RNA samples were isolated from the leaves of all selected *A. thaliana* variants, and RNA-seq libraries were prepared using total RNA depleted of rRNA. For each plant, RNA-seq experiments were performed in four biological replicates (Materials and Methods). On average, 60,707,438 paired-end reads were obtained for the library; 99.99% of the reads passed trimming and quality filtration, 95.95% of them mapped to the *A. thaliana* reference genome (TAIR10), and 7.81% of the reads mapped to the rRNA sequences (Table S2). This observation was consistent with the capillary electrophoresis results (Figure S1).

For circRNA identification, we used the protocol proposed by Gao et al. (CIRI2 [20]) as the best-performing de novo identifying program [28,29]. In total, we identified 30,923 unique circRNAs in Col-0 and all mutants. For each *A. thaliana* variant, we observed that a large fraction of circRNAs (from 84.54% in *cbp80* up to 90.72% in the *tri-20* mutant) was generated in only one biological replicate. CircRNAs that were identified in all four biological replicates were rare (from 0.91% in *tri-20* to 2.62% in *cbp80*) (Figure S2; Table S3A).

The results generated by CIRI2 were additionally validated with *find_circ*, one of the most popular programs for circRNA identification [9]. The outcomes of both analyses were consistent (Figure S2A,B and Table S3B), and with *find_circ* most of the circRNAs were detected in one biological replicate (75.70%) while only a small fraction of the circRNAs was identified in all four replicates (9.60%). In total, 82% of the circRNAs detected by CIRI2 were also found by *find_circ*. Of note, *find_circ* generally identified 1.6 times more circRNAs. This discrepancy might be a result of the different circRNAs identification strategies employed by each program [29], and had no influence on the general conclusion that the majority of circRNAs produced in *A. thaliana* are generated stochastically. To exclude the possibility that the observed phenomenon was caused by the low quality of next-generation sequencing (NGS) data, we checked whether the mRNA accumulation in biological replicates was also variable (see Table S3C,D). In contrast to circRNA, the Pearson correlation coefficients calculated for mRNA were very high, on average $r = 0.9956$ (0.9738–0.9999), while the Pearson correlation coefficients calculated for circRNAs were significantly lower, on average $r = 0.5454$ (from 0.1321 to 0.9773). Thus, the results of the above described mRNA analysis proved that the quality of the NGS data used in our studies was very high. These results are consistent with the observation made in our previous study [30] that most circRNAs in *A. thaliana* are generated stochastically, and thus probably have no biological function. This is why we decided to further focus only on reproducibly generated circRNAs that were present in all four biological replicates.

In total, 24 different circRNAs were identified in the leaves of the wild-type plant. Interestingly, three of these circRNAs (AT1G12240, AT2G44920, AT2G14050) were unique and not present in any other studied variant. All 24 circRNAs were previously reported in PlantcircBase [31], and 16 of them were reported in AtCircDB [15]. Moreover, to validate these results, we checked if the identified circRNAs were also found in a similar analysis that we performed earlier for RNA samples isolated from Col-0 and treated with RNase R prior to library preparation [30]. In total, 22 of them were identified in those libraries. On average, 35 circRNAs were identified per mutant plant (Figure S1). However, we observed that the *cbp80* mutant had a very high number of identified circRNAs (129, Figure 1). The *c2h2* mutant also had a significantly higher-than-average number of circRNAs identified (74). In terms of circRNA abundance, three mutants, namely *cbp80*, *flk* and *c2h2*, accumulated significantly more circRNAs than the wild-type and remaining mutants (Figure 2). The tendency was upheld for *cbp80* and *c2h2* variants when considering all circRNAs produced by the mutants, including the sporadic circRNAs identified in only one biological replicate (Figure S3). This observation suggests that the mutations affect the total

back-splicing process. Notably, all identified circRNAs had canonical splicing signals at the back-splice site (CIRI2 requires the presence of a canonical splicing site to call a circRNA).

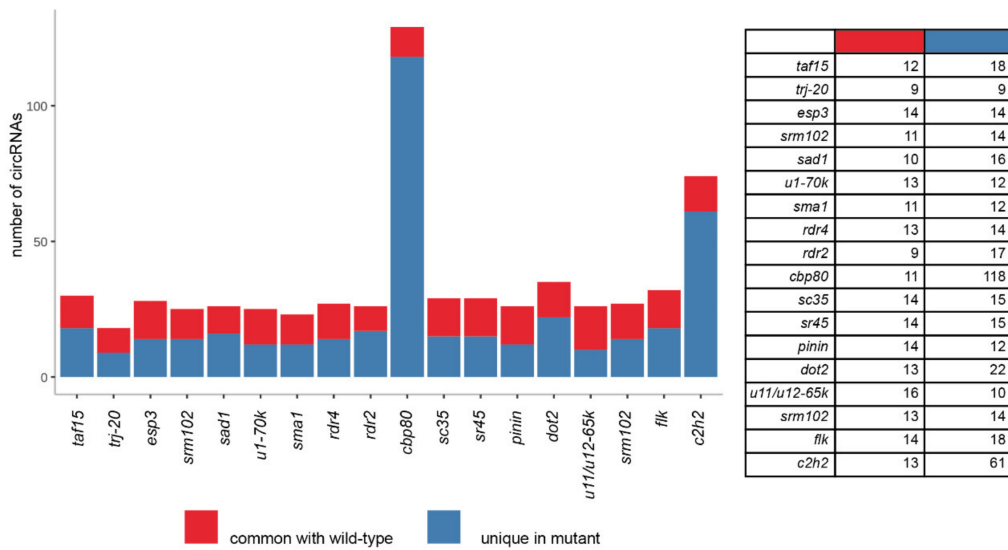


Figure 1. The numbers of circRNAs identified in *A. thaliana* mutants in 4 biological replicates. CircRNAs unique to Col-0 were not shown.

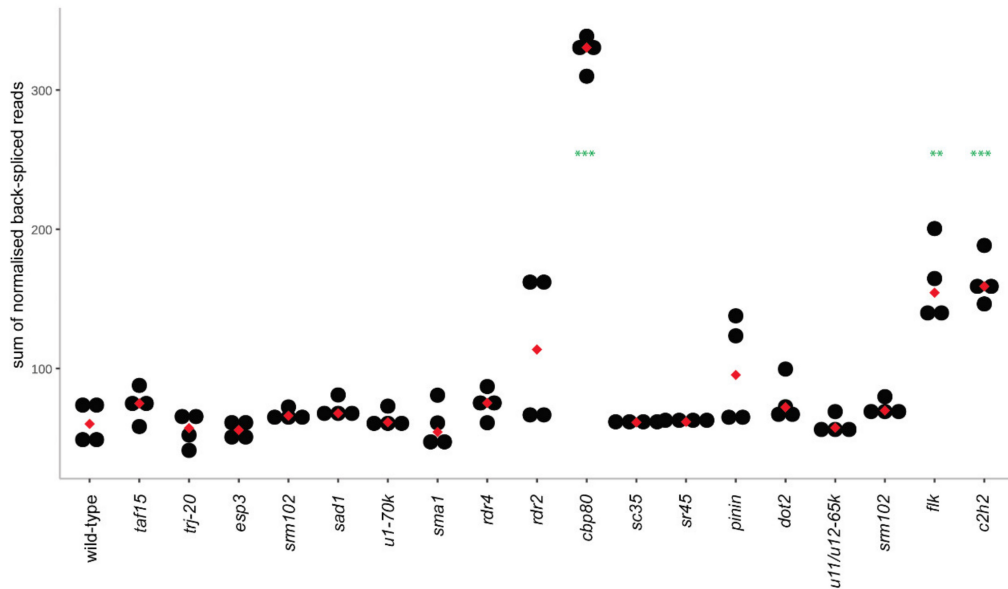


Figure 2. Level of circRNA accumulation in the wild-type and mutant plants [4 biological replicates, ***p*-value ≤ 0.01 , ****p*-value ≤ 0.001).

3.3. CircRNAs in the *cbp80* Mutant

The landscape of circRNAs in the *cbp80* mutant is much more complex than that in the wild-type plant. Among the 129 detected circRNAs, only 11 were common to Col-0, and 26 had at least one mutant (Figure 1, Figure S4). A total of 118 circRNAs seem to be typical to only this mutant, compared with the wild-type. Only 43.2% (51) and 19.5% (23) were previously reported in databases (*PlantcircBase* and *AtCircDB*, respectively).

First, to validate the above results, we verified which of the 118 circRNAs were exclusively specific to *cbp80*. Alternatively, the circRNAs could have been present in Col-0 but in very low amounts, thereby not passing the threshold, as they were not identified in all 4 biological replicates. In fact, 87 circRNAs present in the *cbp80* mutant were unique, and were detected in none of the Col-0 biological replicates. Despite the criteria used, gene ontology (GO) enrichment analyses revealed that genes producing *cbp80*-specific circRNAs were enriched in response to stress (response to cold temperature; p -value < 0.001; Table S4A,B). The remaining 53 circRNAs were common to Col-0. Among these circRNAs, the accumulation of two circRNAs was reduced in the *cbp80* mutant, and that of 8 was increased (Figure 3A; p -value ≤ 0.05 and equal to or greater than a two-fold change).

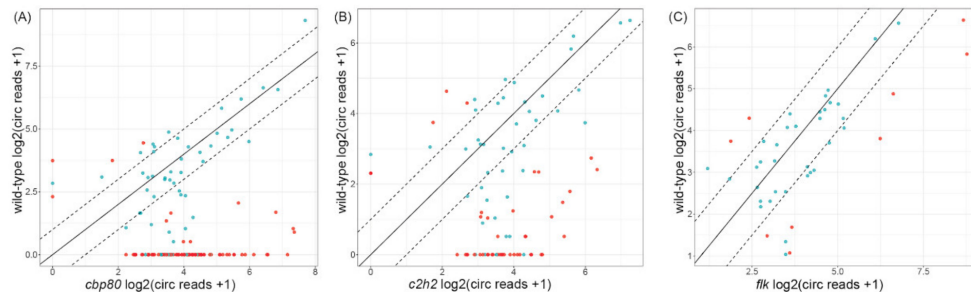


Figure 3. Pairwise comparison of circRNA accumulation in the wild-type plant and (A) *cbp80*, (B) *c2h2* and (C) *flk* mutants. Dashed line: 2-fold cut-off; circRNAs with p -value ≤ 0.05 are marked in red.

We also examined the exon composition of circRNAs from Col-0 and the *cbp80* mutant, and we noticed that most circRNAs identified in Col-0 started from the second or consecutive exons, and those in the *cbp80* mutant started from the first exon (Figure 4). Similar observations were made when the criterion for circRNA identification was relaxed to include circRNAs present in at least 3 biological replicates (see Figure S5). Interestingly, none of the circRNAs identified in the *cbp80* mutant that included the first exon were found in the other 2 mutants that showed increased circRNA accumulation in comparison with the wild-type plant. Moreover, the structure of circRNAs present in the wild-type (Figure S6A) differed significantly from the one observed for circRNAs unique to *cbp80* (Figure S6B). In total 84.5% (109) of *cbp80* circRNAs started or ended in a known exon, whereas in Col-0 56% (14). This observation suggests that alternative splice sites were preferred in the mutant.

Finally, we studied the gene expression profile in the *cbp80* mutant and found that the global gene transcript abundance was not increased compared with that observed for Col-0 (Figure S7). Only 8 genes giving rise to circRNAs in this mutant were downregulated, and 16 were upregulated (Table S5A). No statistically significant GO terms were identified for those differentially expressed genes. Last, we checked the global abundance of the linear counterparts of circRNAs in the *cbp80* mutant, and we observed a statistically significant increase in their abundance (p -value ≤ 0.05 , Figure S8). However, the Pearson correlation coefficient between circular and linear transcripts was 0.015504, indicating that the production of circRNAs was independent of the production of their linear counterparts. The above-mentioned results are clearly reflected in the gene coverage of reads mapped to these genes. There are significant differences in the mapped read number for exons that span circRNAs with increased abundance in the mutant. In Figure 5, we show six examples of genes producing circRNAs with the following features: (i) gene expression level was not changed, but the production of circRNA was significantly increased in the *cbp80* mutant (Figure 5A); (ii) gene expression and circRNA accumulation level was significantly increased in the *cbp80* mutant (Figure 5B); (iii) gene expression and circRNA accumulation level was not changed (Figure 5C). Of note, two of them (namely 3:11289661-11291634 and 5:8188571-8188956) were examples of non-canonical back-splicing occurring in the middle of known exons/introns.

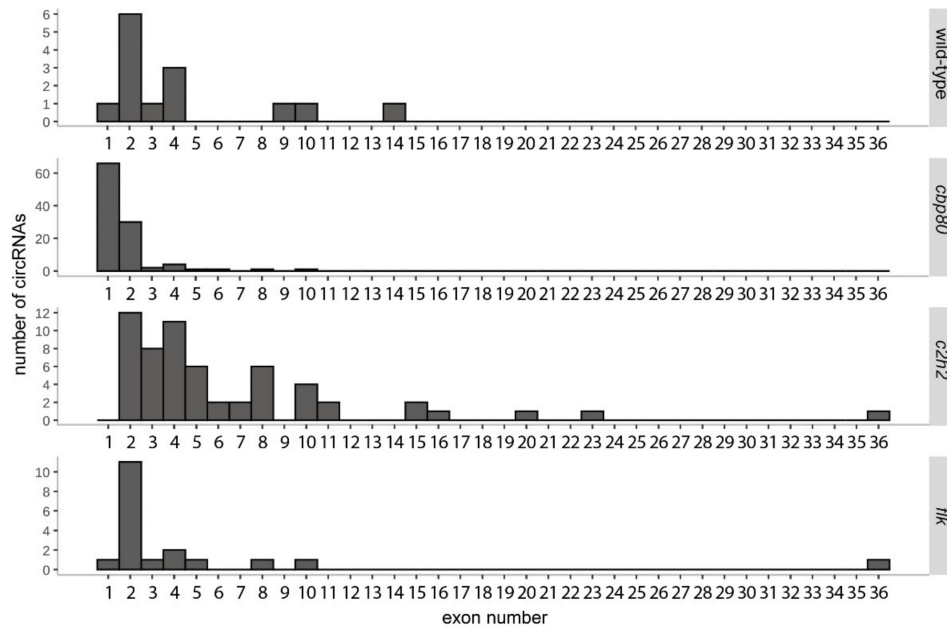


Figure 4. Distribution of exons in which circRNAs start in the wild-type plants and *cbp80*, *c2h2*, and *flk* mutants.

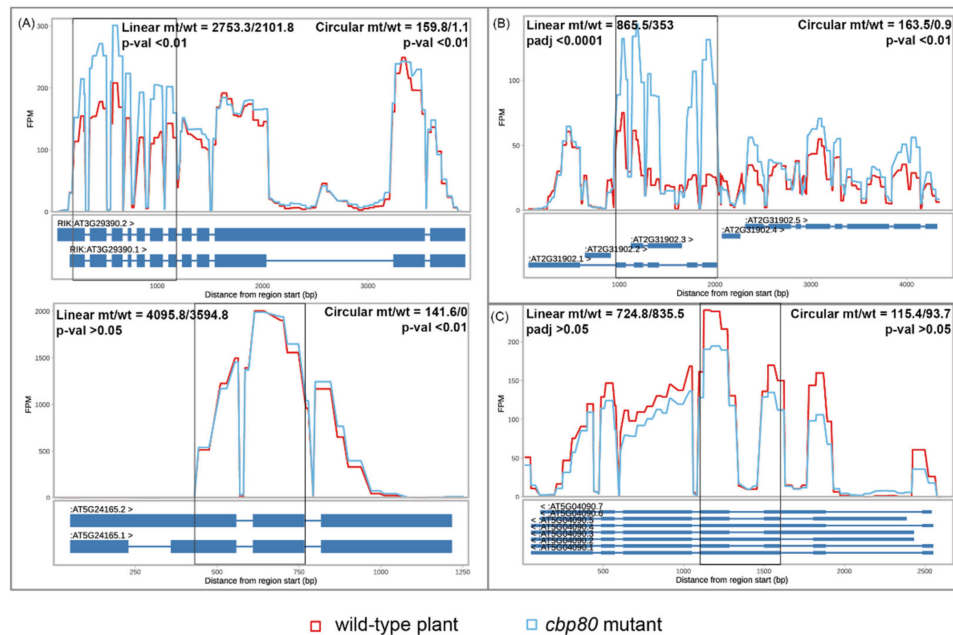


Figure 5. Coverage of genes* giving rise to circRNAs when (A) accumulation of circRNA was increased but parent gene expression was not changed (AT3G29390, 3:11289661-11291634—upper panel; AT5G24165, 5:8188571-8188956—bottom panel), (B) accumulations of circRNA and parent gene expression were increased (AT2G31902, 2:13561602-13563811), and (C) accumulations of circRNA and parent gene expression were not changed (AT5G04090, 5:1106879-1107381). CircRNA ranges are shown in black rectangles. The change of linear RNA accumulation is expressed by the fold change and *p*-value in the left-top of each chart. The change in the circRNA accumulation level is shown by the fold change and *p*-value in the right-top of each chart. * the reads corresponding to both linear and circular RNAs were taken into account.

3.4. CircRNAs in the *c2h2* Mutant

We identified 74 circRNAs in the *c2h2* mutant, which is three times more than the number of circRNAs in the wild-type plant. Among these circRNAs, 13 were common to Col-0 (Figure 1). Sixty-one circRNAs seem to be typical to only this mutant, compared with the wild-type (Figure S4). GO analyses revealed that the genes giving rise to these unique circRNAs were significantly enriched in protein transport and localization (p -value < 0.05, Table S4C). Totals of 80.3% (49) and 54.1% (33) of the circRNAs were previously reported in PlantcircBase and AtCircDB, respectively.

Again, to validate the results, we checked whether the unique circRNAs were present in Col-0 but in very low amounts, and thus might not have passed the threshold for being identified in all four biological replicates. Twenty-eight circRNAs present in the *c2h2* mutant were unique and detected in none of the Col-0 biological replicates. GO analyses revealed that genes giving rise to these unique circRNAs were significantly enriched in cellular processes (p -value < 0.0001; Table S4D). Among the 55 circRNAs found in both the mutant and Col-0, 3 showed decreased accumulation in the mutant and 15 showed increased accumulation (Figure 3B; p -value \leq 0.05 and greater than or equal to a two-fold change). No statistically significant enrichment of genes giving rise to these circRNAs was observed in GO analyses.

CircRNAs from the *c2h2* mutant most frequently started in the second or later exon, as was observed for Col-0, regardless of the circRNA identification criterion used (circRNAs identified in four of the four biological replicates, or in three of four). (Figure 4, see Figure S5). Moreover, the majority of these circRNAs (62, 83.8%) started or terminated within a known exon/intron (Figure S6C).

The global abundance of transcripts in the *c2h2* mutant was not increased (Figure S7). The expressions of only three genes producing circRNAs were downregulated (AT1G60590, AT3G46970, AT2G21320; Table S5B). Again, we observed no correlation between circular and linear transcript production (Pearson correlation coefficient of -0.075489266). Moreover, the abundance of the linear counterparts of the circRNAs was increased in the *c2h2* mutant (p -value \leq 0.05, Figure S8).

3.5. CircRNAs in the *flk* Mutant

In total, 32 circRNAs were identified in the *flk* mutant. All of these circRNAs were also present in Col-0; however, 18 circRNAs did not pass the criterion of being identified in all 4 biological replicates of Col-0 (Figure 1). Among the *flk* circRNAs, 94.4% (17) and 66.7% (12) were previously reported in PlantcircBase and AtCircDB, respectively. In contrast to the *cbp80* and *c2h2* mutants, no circRNAs specific to this mutant were identified (Figure S4). The analyses of circRNA abundance revealed that the accumulation of seven circRNAs was increased, and that of two was decreased in comparison to their abundance in Col-0 (Figure 3C; p -value \leq 0.05 and a greater than two-fold change). No statistically significant GO enrichment was observed.

Most circRNAs in the *flk* mutant started in the second exon, as in Col-0 (Figure 4). Similar observations were made when the criterion for circRNA identification was relaxed to include circRNAs present in at least three biological replicates (see Figure S5). Similarly to Col-0, 62.5% (20) of the circRNAs started or ended within a known exon/intron (Figure S6D).

The general abundance of transcripts in the *flk* mutant was not increased (Figure S7), and no correlation between circular and linear counterparts was observed (Pearson correlation coefficient of -0.05346). Only one gene (AT4G30975) giving rise to circRNAs in the *flk* mutant was upregulated (Table S5C). On average, the abundance of the linear counterparts of circular transcripts was increased (p -value \leq 0.001, Figure S8).

3.6. CircRNA Production as a Form of Alternative Splicing

Considering that splicing machinery is involved in circRNAs biogenesis, we checked whether the production of alternatively spliced linear transcripts was also altered in the analyzed *A. thaliana* variants. The analysis of all splicing variants identified in the studied plants revealed that on average,

8.9% of the transcripts (standard deviation: st-dev. 3.7%) per mutant displayed significant differences in the levels of expression compared to the wild-type (Figure S9). Mutant *u11/u12-65k* was characterized by a very low number of alternatively spliced transcripts (as well as a very low number of circRNAs, specifically, 10 if circRNAs identified in four biological replicates in the wild-type were considered and 0 if all circRNAs in the wild-type were considered), suggesting that the mutation did not significantly affect splicing. The three mutants *rdr4*, *cbp80* and *c2h2* had respectively 1.6-, 2.2- and 1.7-fold more alternatively spliced transcripts than the average value determined for all 18 mutants. Interestingly, we identified *cbp80* and *c2h2* mutants as producing higher amounts and numbers of circRNAs in comparison to Col-0 (Figure 1).

We analyzed the gene structure of the top 10 alternatively spliced transcripts, and noticed that mutant plants generated forms of transcripts other than Col-0 for these genes (Figure S10A,B). When only genes producing circRNAs (identified in all four biological replicates) were considered, we again noticed significant alternative splicing events in the *cbp80* and *c2h2* variants (Figure S11). In *cbp80*, 38 transcripts were identified as up- (20) or down- (18) regulated, and in *c2h2*, 3 transcripts were significantly upregulated and 6 downregulated. Moreover, in these two mutants, in contrast to other analyzed variants, we found transcripts (21 in *cbp80* and 8 in *c2h2*) the level of which displayed small (less than two-fold) but statistically significant changes in comparison to the wild-type plant ($p_{adj} \leq 0.05$, Figure S11). We checked whether there were exon skipping events that could be coupled to back-splicing (where the skipped exon forms a circRNA), however we found no evidence of such a phenomenon.

4. Discussion

CircRNAs have been identified in multiple plant species under stress conditions (reviewed in [32]). Several studies have provided evidence that circRNA pools are dynamic and fluctuate in response to external stimuli. Based on these observations, circRNAs have been proposed to possess biological functions [5–7]. Although very intriguing, until now, such hypotheses have been supported by only one direct experimental proof of the functional relevance of circRNA in plants [16]. Considering that circRNAs do play important roles in animal and human cells, evidence of their functions in plants is expected.

CircRNA biogenesis in plants is associated with splicing; however, the molecular mechanism of circRNA formation has not been fully elucidated thus far. To better understand this phenomenon, we analyzed 18 *A. thaliana* knockout mutants for genes that encode proteins involved in different stages of splicing. We demonstrated previously [30], as well as in this study, that only a very limited fraction of circRNAs (less than 3%) are generated in a reproducible manner, and thus can serve as functional molecules. Consequently, the majority of circRNAs most likely reflect the imperfections of splicing. This observation raises questions about the mechanisms that prevent circRNA interference with RNA metabolism, RNA–protein interactions, and RNA-dependent regulatory pathways. Further studies of circRNA intracellular localization are required in order to track the fates of circRNAs and dissect the appropriate surveillance systems.

The production of circRNAs in three mutants was significantly different from that observed in the wild-type and other variants. The first mutant lacked the functional *cbp80* gene, the product of which is a component of the cap-binding complex. The second mutant, *c2h2*, lacked a protein of the U4/U6.U5 tri-snRNP pre-assembled spliceosomal complex that plays a key role in the formation of a catalytically active spliceosome. The third mutant, *flk*, lacked a member of the group of nuclear ribonucleoproteins (hnRNPs), which, if bound to pre-mRNA molecules, serve as a signal that the pre-mRNA is not yet fully processed, and therefore not ready for export to the cytoplasm. The common feature of these three mutants was the increased accumulation of circular transcripts (compared to that in the wild-type plant), which was not accompanied by corresponding changes in gene expression. The number of alternative splicing events was increased in these mutants. The observed higher accumulation of circRNA could be due to either more frequent back-splicing or increased circRNA stability in mutants.

This problem requires further detailed studies, and one has to acknowledge that both mechanisms can shape the circRNA pool. However, the fact that we did detect increased numbers of alternatively spliced linear transcripts in mutants indicates that the splicing itself was affected by the gene knockout (currently there is no evidence that these linear alternatively spliced variants could display higher stability than their canonically spliced counterparts). Altogether, these findings suggest that the expression of linear transcripts might be regulated through circRNA production. In the majority of the analyzed mutants, the depletion of splicing-related proteins did not have a significant effect on the generation of circRNAs. This observation indicates that there exist only defined checkpoints prone to either disturbance or regulation, which can change the balance between linear and circular transcripts.

Among the 18 plant variants tested in this study, *cbp80* was found to produce the largest number of unique circRNAs. CBP80 together with CBP20 forms a cap-binding complex (CBC), implicated in splicing and miRNA biogenesis. *A. thaliana* mutants lacking CBC or its components accumulate partially spliced transcripts [33] and pri-miRNA [34]. In addition, the lack of CBC and, in particular, CBP80 affects alternative splicing, especially the selection of the 5' splice site of the first intron [35]. Our data provide, for the first time, evidence that the absence of CBP80 not only disturbs the production of linear transcript isoforms, but also dramatically increases the generation of circRNAs. Notably, the majority of *cbp80*-specific circRNAs included the first exon, while those formed in other mutants and Col-0 were produced mostly from the second or consecutive exons. This observation is consistent with previous findings that CBP80 preferentially exerts its effect on splicing in the mRNA region proximal to the 5' end.

CBP80 is involved in abscisic acid (ABA) signal transduction [36] and the flowering pathway [37,38]. Mutants lacking CBP80 display an ABA-induced elevation of cytosolic calcium levels in guard cells, leading to enhanced stomatal closure and the provision of increased drought tolerance [36]. These plants also show an early-flowering phenotype, resulting from the defective splicing of FLOWERING LOCUS C (FLC) introns, which decreases the level of properly spliced FLC transcripts [37]. We wondered whether the generation of circRNAs in *cbp80* also contributed to the observed phenotypes; however, our preliminary analyses did not provide any clues. The repertoire of *cbp80*-specific circRNAs has to be examined in greater detail in order to provide definitive conclusions regarding this aspect.

CircRNAs can either be generated in a controlled way or constitute products of mis-splicing. Based on our data, one can hypothesize that, in general, the undisturbed functioning of protein factors involved in splicing favors the formation of linear transcripts. Such a model appears consistent with the cellular RNA metabolism pathways, which to a large extent rely upon exonucleases, implicated both in quality control processes and in the regulation of transcript stability. Nevertheless, our results suggest that a remodeling of the splicing-related protein complexes may provide a facile mechanism to trigger a switch toward circRNA production. The dynamic composition of the cap-binding protein complex in *A. thaliana* is well documented, and changes during the growth cycle [39]. CBP80 is among the proteins that remain associated with the mRNA cap in both proliferating and quiescent cells, but it changes subcellular localization. During the early stages of the growth cycle, this protein preferentially localizes to the cytoplasm, while later, it accumulates in the nucleus. In turn, its partner, CBP20, first localizes to both the nucleus and the cytoplasm, and then is concentrated in the nucleus [39]. One can hypothesize that some splicing events early in the growth cycle may occur under CBP80 deficiency. Such a situation, according to our results, may activate the formation of circRNAs. Therefore, the recruitment of diverse splicing-related factors (for example, factors dependent on the strength of splicing signals or enhancers) or the sequestration of the components of splicing machinery emerge as potential mechanisms of gene expression regulation in plants. Interestingly, the previous observation that circRNA expression increases when core spliceosomal components are depleted in *Drosophila* [40] suggests the conservation of such a mechanism across kingdoms. Detailed studies are required to provide further insight into this issue in *A. thaliana*.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4409/9/9/1937/s1>, Figure S1: The efficiency of Ribo-Zero treatment, Figure S2: Number of identified circRNAs by (A) CIRI2, and (B) find_circ in 1,2,3, or 4 biological replicates, Figure S3: Level of circRNA accumulation in the wild-type and mutant plants, Figure S4: Number of common and unique circRNAs identified in the wild-type and mutant plants, Figure S5: Number of circRNAs, identified in minimum 3 biological replicates, starting with the specified exon number in the wild-type plant, *cbp80*, *c2h2*, and *flk* mutants, Figure S6: The occurrence of canonical and non-canonical back-splicing sites in the wild-type, *cbp80*, *c2h2*, and *flk* mutants, Figure S7: Level of gene expression in the wild-type plant and *cbp80*, *c2h2*, *flk* mutants, Figure S8: Accumulation level of circRNAs' linear counterparts in the wild-type and mutant plants, Figure S9: The percentage of differentially expressed transcripts in mutants compared to the wild-type plant, Figure S10: The percentage of differentially expressed transcripts in mutants compared to the wild-type plant, Figure S11: The expression profile of transcripts from genes generating circRNAs (identified in 4 biological replicates) in mutants vs. wild-type plants. Table S1: List of primers used in the study, Table S2: The information for the RNA-seq libraries, Table S3: A) CircRNAs identified with CIRI2; B) CircRNAs identified with find_circ; C) Number of normalized reads for linear transcripts; D) Pearson correlation coefficient of mRNA reads between biological replicates, Table S4: Gene ontology for genes giving rise to circRNA, Table S5: A) List of circRNA identified in *cbp80* in 4 biological replicates; B) List of circRNA identified in *c2h2* in 4 biological replicates; C) List of circRNA identified in *flk* in 4 biological replicates. The datasets supporting the conclusions of this article are available in the National Center for Biotechnology Information Sequence Read Archive repository.

Author Contributions: M.F. conceived the overall idea of the study. A.P., P.J. and M.F. conceived and designed the experiments. A.P., P.J. and M.F. analyzed the data and discussed the results. K.N. performed all bioinformatics analyses. M.S. cultivated plants and isolated RNA. M.S. and J.P. prepared RNA samples for sequencing. J.P. and L.H. generated the RNA-seq libraries. A.P., K.N., M.S. and P.J. drafted the manuscript. M.F. was responsible for the final version of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financed by the National Science Centre grant number UMO-2014/15/D/NZ2/02305 to AP. Partial financial support was also provided by the Polish Ministry of Science and Higher Education under the KNOW program.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, P.L.; Bao, Y.; Yee, M.-C.; Barrett, S.P.; Hogan, G.J.; Olsen, M.N.; Dinneny, J.R.; Brown, P.O.; Salzman, J. Circular RNA Is Expressed across the Eukaryotic Tree of Life. *PLoS ONE* **2014**, *9*, e90859. [[CrossRef](#)] [[PubMed](#)]
2. Jeck, W.R.; Sorrentino, J.A.; Wang, K.; Slevin, M.K.; Burd, C.E.; Liu, J.; Marzluff, W.F.; Sharpless, N.E. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **2012**, *19*, 141–157. [[CrossRef](#)] [[PubMed](#)]
3. Rybak-Wolf, A.; Stottmeister, C.; Glažar, P.; Jens, M.; Pino, N.; Giusti, S.; Hanan, M.; Behm, M.; Bartok, O.; Ashwal-Fluss, R.; et al. Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol. Cell* **2015**, *58*, 870–885. [[CrossRef](#)] [[PubMed](#)]
4. Lasda, E.; Parker, R. Circular RNAs: diversity of form and function. *RNA* **2014**, *20*, 1829–1842. [[CrossRef](#)] [[PubMed](#)]
5. Qu, S.; Liu, Z.; Yang, X.; Zhou, J.; Yu, H.; Zhang, R.; Li, H. The emerging functions and roles of circular RNAs in cancer. *Cancer Lett.* **2018**, *414*, 301–309. [[CrossRef](#)] [[PubMed](#)]
6. Hansen, T.B.; I Jensen, T.; Clausen, B.H.; Bramsen, J.B.; Finsen, B.; Damgaard, C.K.; Kjems, J. Natural RNA circles function as efficient microRNA sponges. *Nature* **2013**, *495*, 384–388. [[CrossRef](#)]
7. Ashwal-Fluss, R.; Meyer, M.; Pamudurti, N.R.; Ivanov, A.; Bartok, O.; Hanan, M.; Evantal, N.; Memczak, S.; Rajewsky, N.; Kadener, S. circRNA Biogenesis Competes with Pre-mRNA Splicing. *Mol. Cell* **2014**, *56*, 55–66. [[CrossRef](#)]
8. Du, W.W.; Zhang, C.; Yang, W.; Yong, T.; Awan, F.M.; Yang, B.B. Identifying and Characterizing circRNA-Protein Interaction. *Theranostics* **2017**, *7*, 4183–4191. [[CrossRef](#)]
9. Memczak, S.; Jens, M.; Elefsinioti, A.; Torti, F.; Krueger, J.; Rybak, A.; Maier, L.; Mackowiak, S.D.; Gregersen, L.H.; Munschauer, M.; et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **2013**, *495*, 333–338. [[CrossRef](#)]
10. Peng, Y.; Song, X.; Zheng, Y.; Cheng, H.; Lai, W. circCOL3A1-859267 regulates type I collagen expression by sponging miR-29c in human dermal fibroblasts. *Eur. J. Dermatol. EJD* **2018**, *28*, 613–620.
11. Liu, J.; Li, D.; Luo, H.; Zhu, X. Circular RNAs: The star molecules in cancer. *Mol. Asp. Med.* **2019**, *70*, 141–152. [[CrossRef](#)] [[PubMed](#)]

12. Patop, I.L.; Wüst, S.; Kadener, S. Past, present, and future of circ RNA s. *EMBO J.* **2019**, *38*. [[CrossRef](#)] [[PubMed](#)]
13. Lukiw, W.J. Circular RNA (circRNA) in Alzheimer's disease (AD). *Front. Genet.* **2013**, *4*. [[CrossRef](#)] [[PubMed](#)]
14. Liang, D.; E Wilusz, J. Short intronic repeat sequences facilitate circular RNA production. *Genes Dev.* **2014**, *28*, 2233–2247. [[CrossRef](#)] [[PubMed](#)]
15. Ye, J.; Wang, L.; Li, S.; Zhang, Q.; Zhang, Q.; Tang, W.; Wang, K.; Song, K.; Sablok, G.; Sun, X.; et al. AtCircDB: a tissue-specific database for *Arabidopsis* circular RNAs. *Briefings Bioinform.* **2017**, *20*, 58–65. [[CrossRef](#)] [[PubMed](#)]
16. Conn, V.M.; Hugouvieux, V.; Nayak, A.; Conos, S.A.; Capovilla, G.; Cildir, G.; Jourdain, A.; Tergaonkar, V.; Schmid, M.; Zubieta, C.; et al. A circRNA from SEPALLATA3 regulates splicing of its cognate mRNA through R-loop formation. *Nat. Plants* **2017**, *3*, 17053. [[CrossRef](#)]
17. Wang, B.-B.; Brendel, V. The ASRG database: identification and survey of *Arabidopsis thaliana* genes involved in pre-mRNA splicing. *Genome Biol.* **2004**, *5*, R102. [[CrossRef](#)]
18. Boyes, D.C.; Zayed, A.M.; Ascenzi, R.; McCaskill, A.J.; Hoffman, N.E.; Davis, K.R.; Görlach, J. Growth Stage-Based Phenotypic Analysis of *Arabidopsis*. *Plant Cell* **2001**, *13*, 1499–1510. [[CrossRef](#)]
19. Schubert, M.; Lindgreen, S.; Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **2016**, *9*, 88. [[CrossRef](#)]
20. Gao, Y.; Wang, J.; Zhao, F. CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol.* **2015**, *16*, 4. [[CrossRef](#)]
21. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]
22. Kopylova, E.; Noé, L.; Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **2012**, *28*, 3211–3217. [[CrossRef](#)] [[PubMed](#)]
23. Perteza, M.; Kim, D.; Perteza, G.M.; Leek, J.T.; Salzberg, S.L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **2016**, *11*, 1650–1667. [[CrossRef](#)] [[PubMed](#)]
24. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [[CrossRef](#)]
25. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [[CrossRef](#)]
26. I Love, M.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **2014**, *15*, 002832. [[CrossRef](#)]
27. ABRC. Available online: <http://abrc.osu.edu> (accessed on 4 June 2019).
28. Hansen, T.B. Improved circRNA Identification by Combining Prediction Algorithms. *Front. Cell Dev. Biol.* **2018**, *6*. [[CrossRef](#)]
29. Szabo, L.; Salzman, J. Detecting circular RNAs: bioinformatic and experimental challenges. *Nat. Rev. Genet.* **2016**, *17*, 679–692. [[CrossRef](#)]
30. Philips, A.; Nowis, K.; Stelmaszczyk, M.; Jackowiak, P.; Podkowiński, J.; Handschuh, L. Only a small fraction of circular RNAs is reproducibly formed in *Arabidopsis thaliana* seedlings and organs. *BMC Plant Biol.* **2020**. Conditionally Accepted.
31. Chu, Q.; Zhang, X.; Zhu, X.; Liu, C.; Mao, L.; Ye, C.; Zhu, Q.-H.; Fan, L. PlantcircBase: A Database for Plant Circular RNAs. *Mol. Plant* **2017**, *10*, 1126–1128. [[CrossRef](#)]
32. Zhao, W.; Chu, S.; Jiao, Y. Present Scenario of Circular RNAs (circRNAs) in Plants. *Front. Plant Sci.* **2019**, *10*. [[CrossRef](#)] [[PubMed](#)]
33. Laubinger, S.; Sachsenberg, T.; Zeller, G.; Busch, W.; Lohmann, J.U.; Ratsch, G.; Weigel, D. Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 8795–8800. [[CrossRef](#)]
34. Kim, S.; Yang, J.-Y.; Xu, J.; Jang, I.-C.; Prigge, M.J.; Chua, N.-H.; Kurasawa, K.; Matsui, A.; Yokoyama, R.; Kuriyama, T.; et al. Two Cap-Binding Proteins CBP20 and CBP80 are Involved in Processing Primary MicroRNAs. *Plant Cell Physiol.* **2008**, *49*, 1634–1644. [[CrossRef](#)] [[PubMed](#)]
35. Raczynska, K.D.; Simpson, C.G.; Ciesiolka, A.; Szewc, L.; Lewandowska, D.; McNicol, J.; Szweykowska-Kulinska, Z.; Brown, J.W.S.; Jarmolowski, A. Involvement of the nuclear cap-binding protein complex in alternative splicing in *Arabidopsis thaliana*. *Nucleic Acids Res.* **2009**, *38*, 265–278. [[CrossRef](#)] [[PubMed](#)]

36. Hugouvieux, V.; Murata, Y.; Young, J.J.; Kwak, J.M.; Mackesy, D.Z.; I Schroeder, J. Localization, Ion Channel Regulation, and Genetic Interactions during Abscisic Acid Signaling of the Nuclear mRNA Cap-Binding Protein, ABH11. *Plant Physiol.* **2002**, *130*, 1276–1287. [[CrossRef](#)] [[PubMed](#)]
37. Kuhn, J.M.; Breton, G.; I Schroeder, J. mRNA metabolism of flowering-time regulators in wild-type Arabidopsis revealed by a nuclear cap binding protein mutant, abh1. *Plant J.* **2007**, *50*, 1049–1062. [[CrossRef](#)]
38. Kuhn, J.M.; Hugouvieux, V.; I Schroeder, J. mRNA Cap Binding Proteins: Effects on Abscisic Acid Signal Transduction, mRNA Processing, and Microarray Analyses. In *Nuclear pre-mRNA Processing in Plants*; Springer: Berlin, Germany, 2008; pp. 139–150.
39. Bush, M.S.; Hutchins, A.P.; Jones, J.D.; Naldrett, M.J.; Jarmolowski, A.; Lloyd, C.W.; Doonan, J.H. Selective recruitment of proteins to 5' cap complexes during the growth cycle in Arabidopsis. *Plant J.* **2009**, *59*, 400–412. [[CrossRef](#)]
40. Liang, N.; Tatomer, D.C.; Luo, Z.; Wu, H.; Yang, L.; Chen, L.-L.; Cherry, S.; E Wilusz, J. The Output of Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is Limiting. *Mol. Cell* **2017**, *68*, 940–954.e3. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

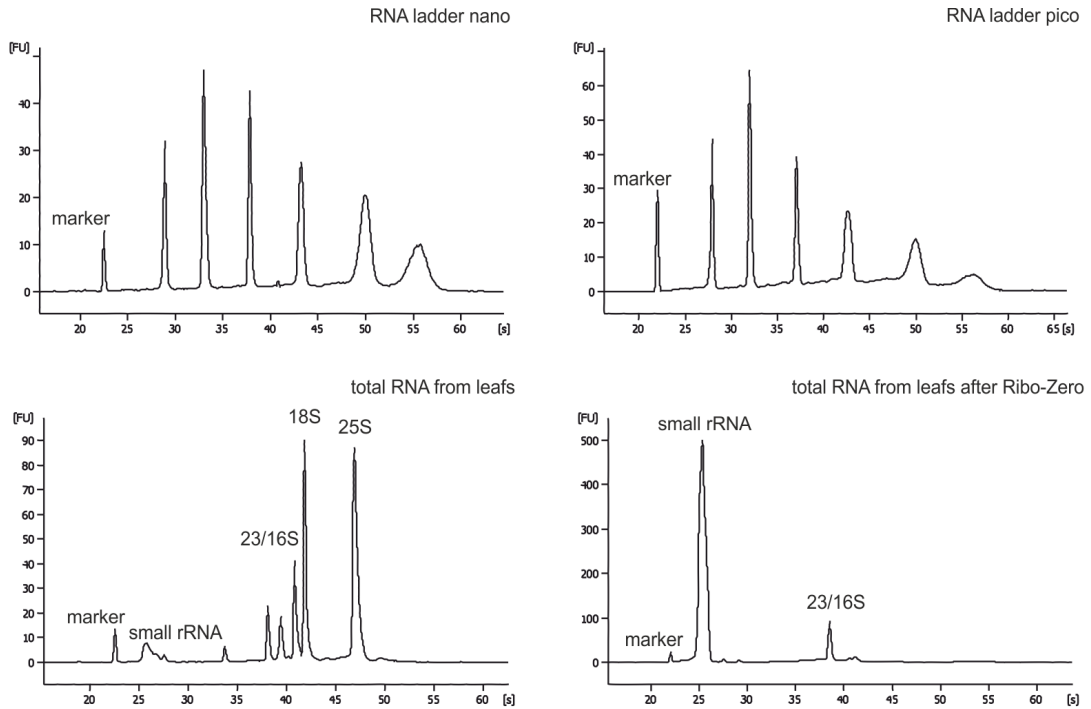
Materiały Suplementarne

Philips A, **Nowis K**, Stelmaszczuk M, Podkowiński J, Handschuh L, Jackowiak P,
Figlerowicz M;

Arabidopsis thaliana cbp80, c2h2, and flk Knockout Mutants Accumulate
Increased Amounts of Circular RNAs

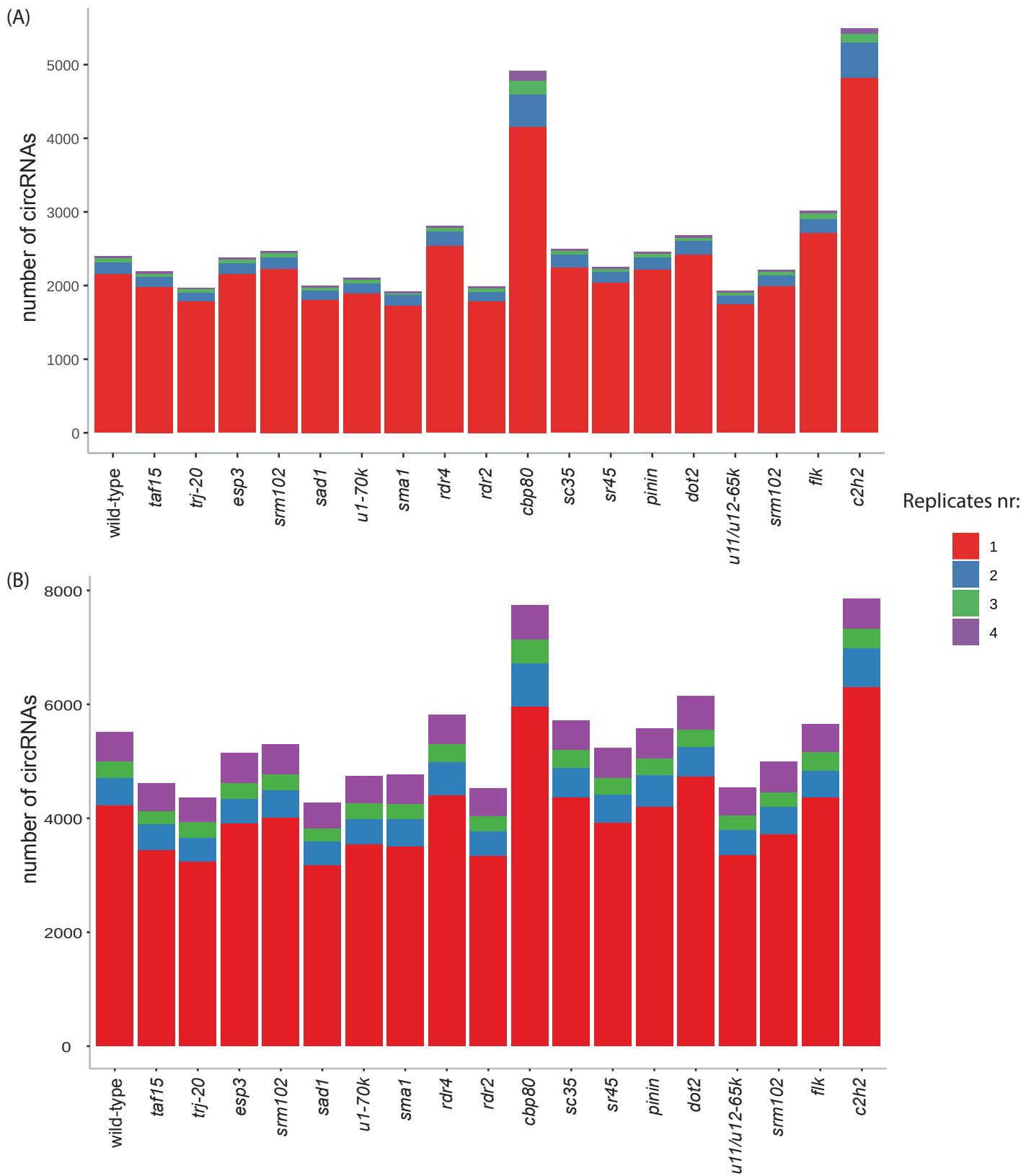
Cells (2020), <https://doi.org/10.3390/cells9091937>

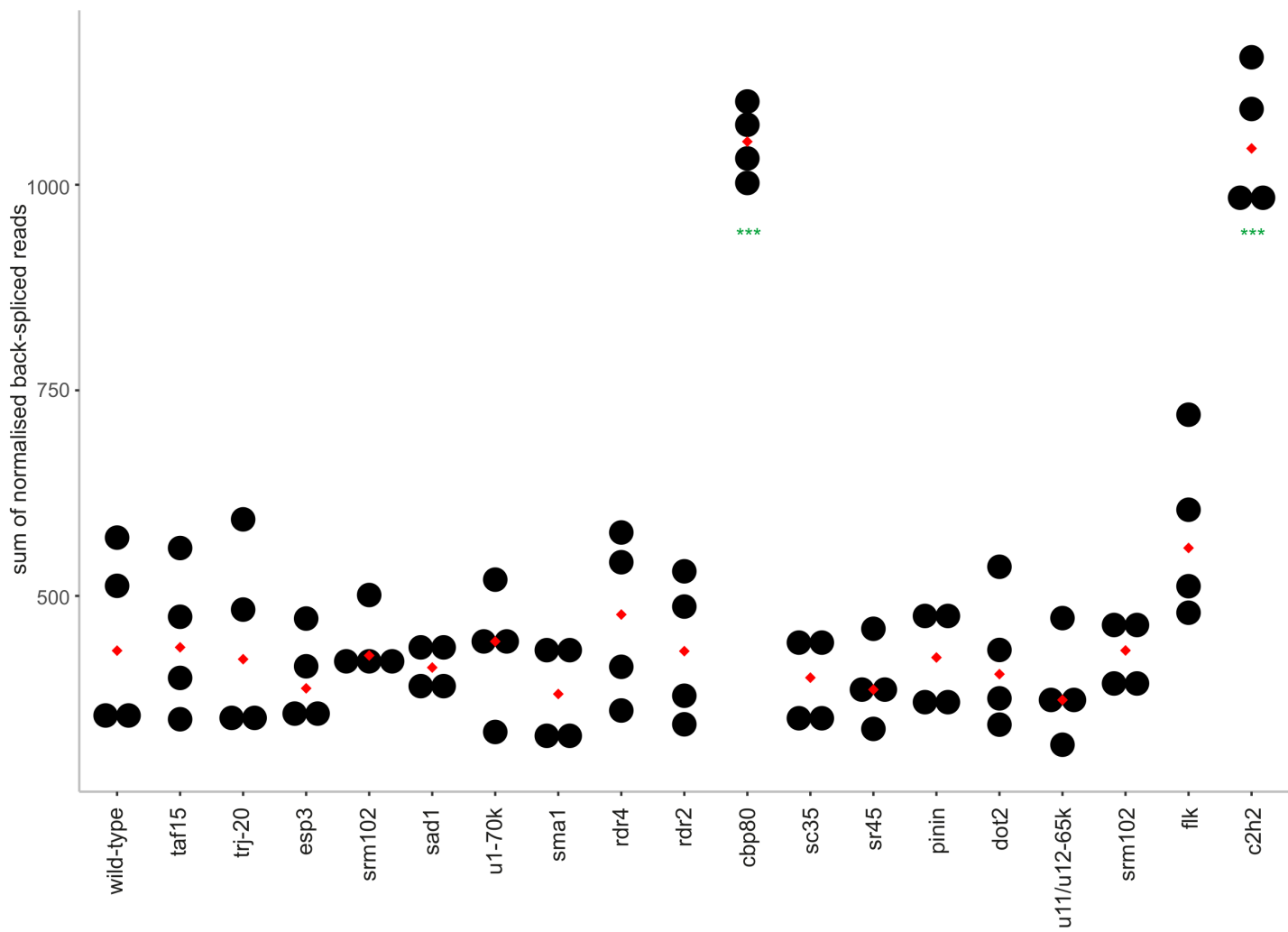
Pięcioletni IF 7,7, MEiN 140



Supplementary Figure 1.
The efficiency of Ribo-Zero treatment.

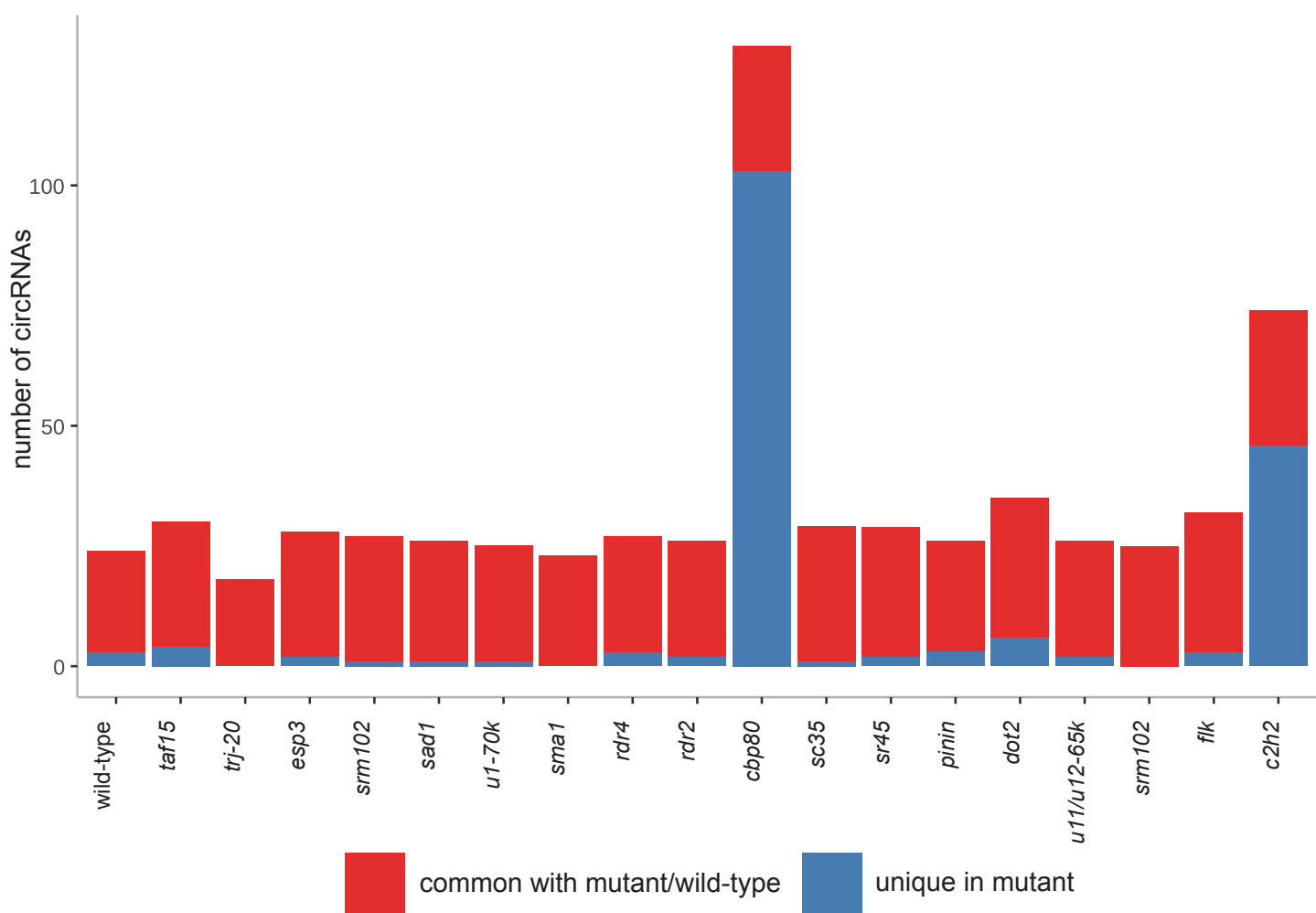
Representative electropherograms (Bioanalyzer, Agilent) of total RNAs (left side), total RNAs after Ribo-Zero treatment (right side) from leaves of *A. thaliana*. One hundred nanograms of DNase-treated total RNA of each sample was analyzed on the Bioanalyzer, Agilent with RNA Nano Chip in accordance with the manufacturer's instructions. After Ribo-Zero treatment, 3 ng of each sample was analyzed on the Bioanalyzer, Agilent with RNA Pico Chip in accordance with the manufacturer's instructions. The x-axis represents the resolving time in seconds (s), and the y-axis represents fluorescence (fu). In addition to the cytoplasmic 25S and 18S rRNA peaks, other peaks corresponding to 23S and 16S rRNA from chloroplasts and small rRNAs are present.





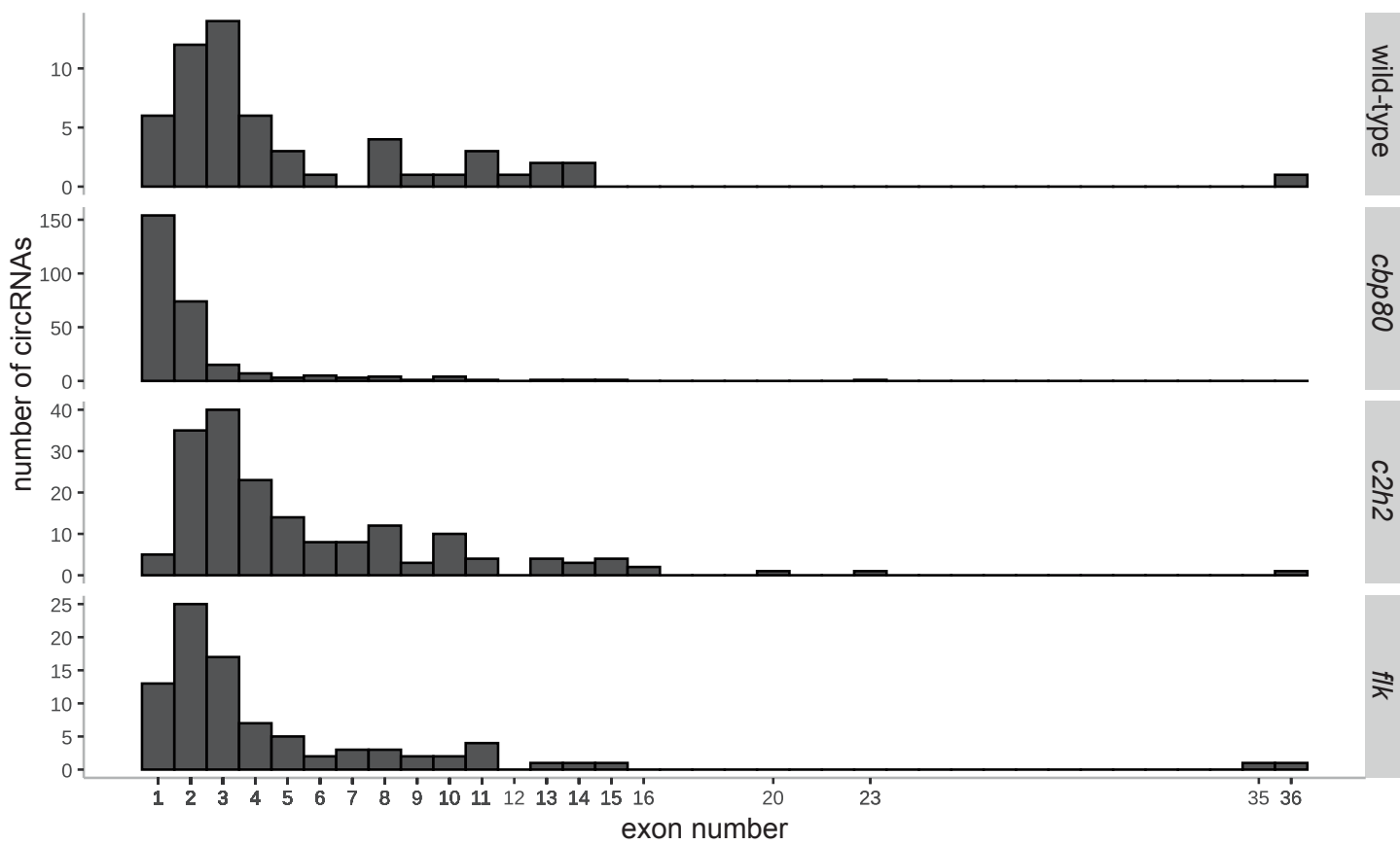
Supplementary Figure 3.

Level of circRNA accumulation in the wild-type and mutant plants (all identified circRNAs, p-value ≤ 0.05 *, ≤ 0.01 **, ≤ 0.001 ***).



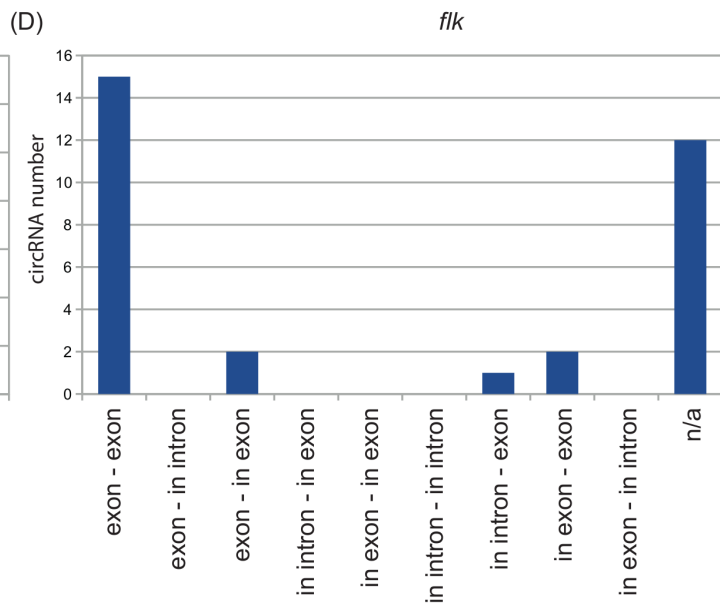
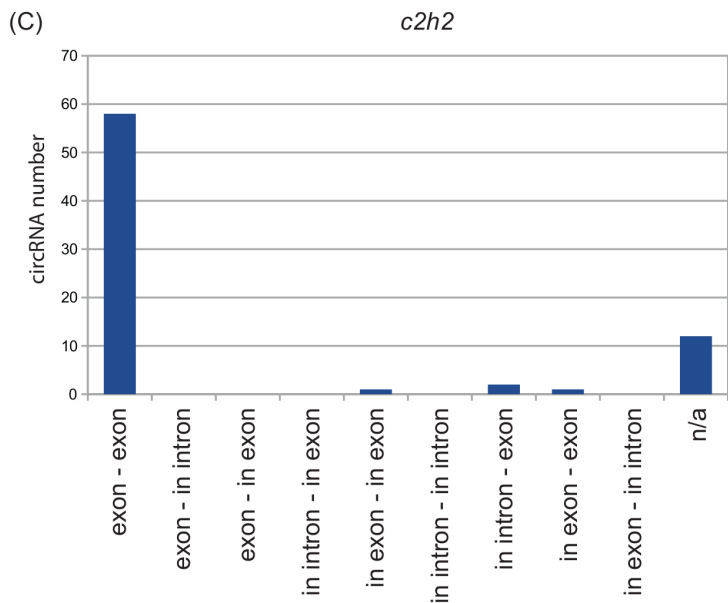
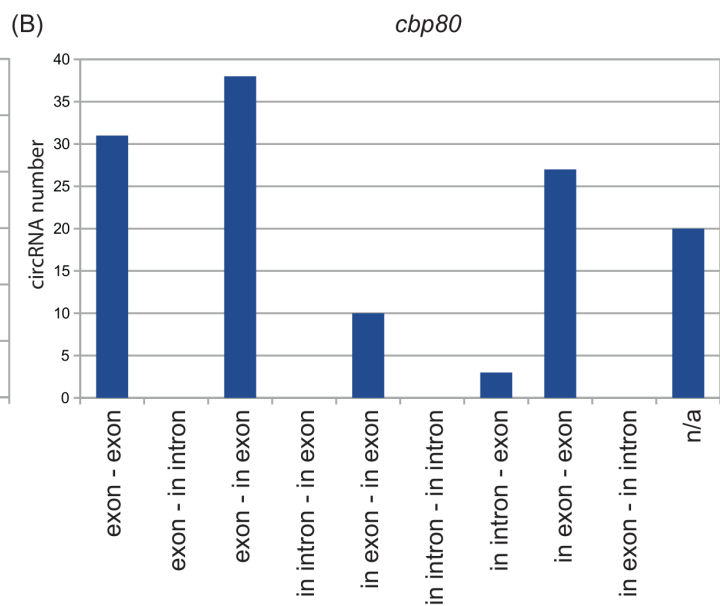
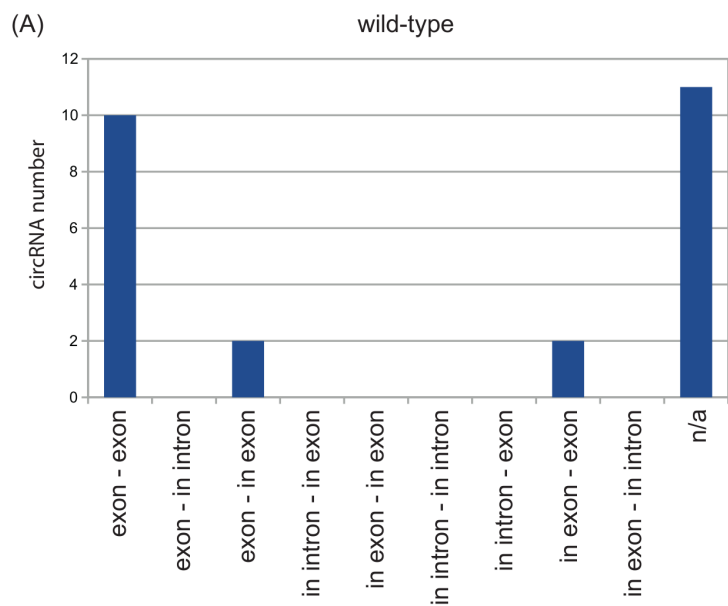
Supplementary Figure 4.

Number of common and unique circRNAs identified in the wild-type and mutant plants.



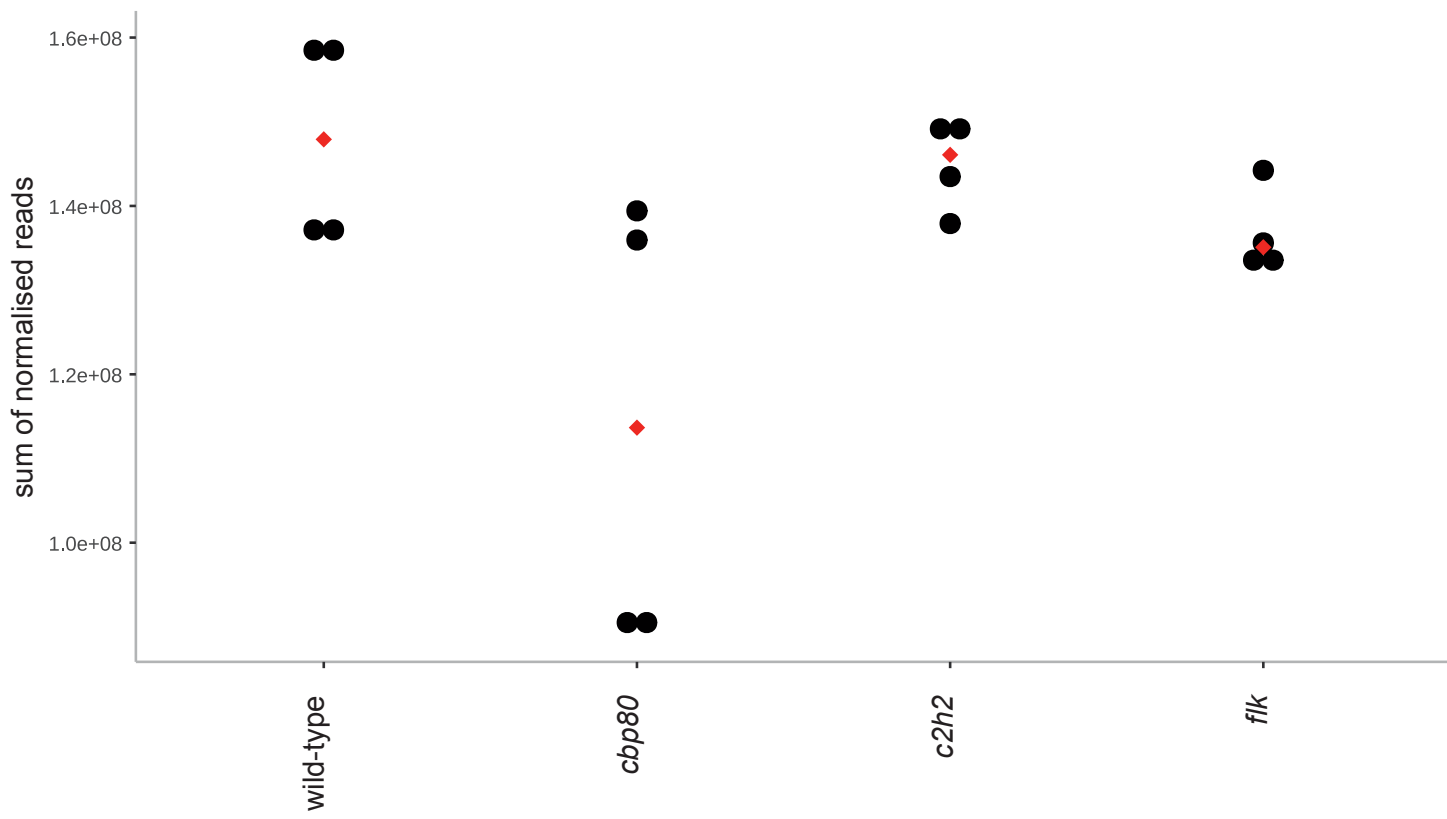
Supplementary Figure 5.

Number of circRNAs, identified in minimum 3 biological replicates, starting with the specified exon number in the wild-type plant, *cbp80*, *c2h2*, and *flk* mutants.



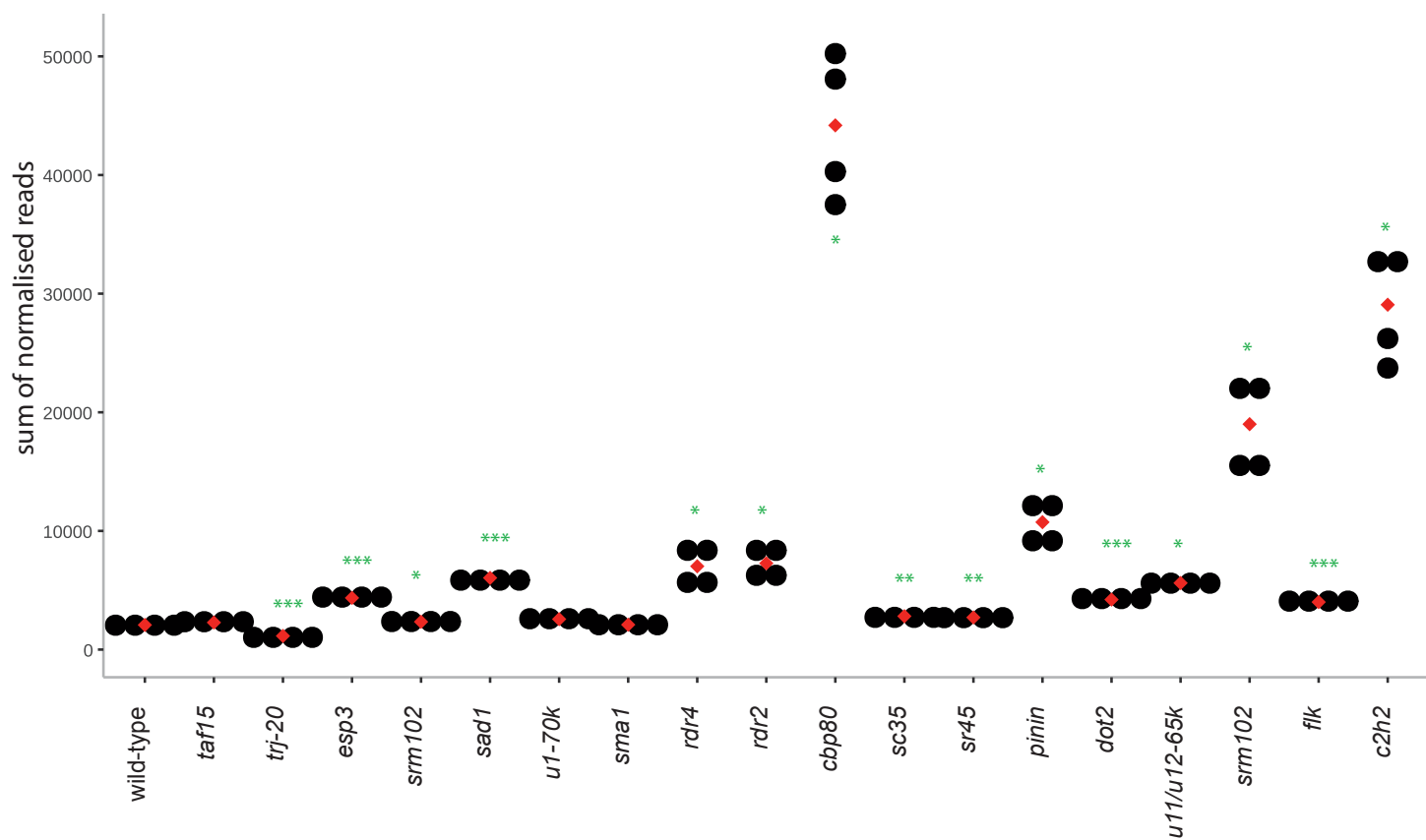
Supplementary Figure 6.

The occurrence of canonical and non-canonical back-splicing sites in the wild-type, *cbp80*, *c2h2*, and *flk* mutants.



Supplementary Figure 7.

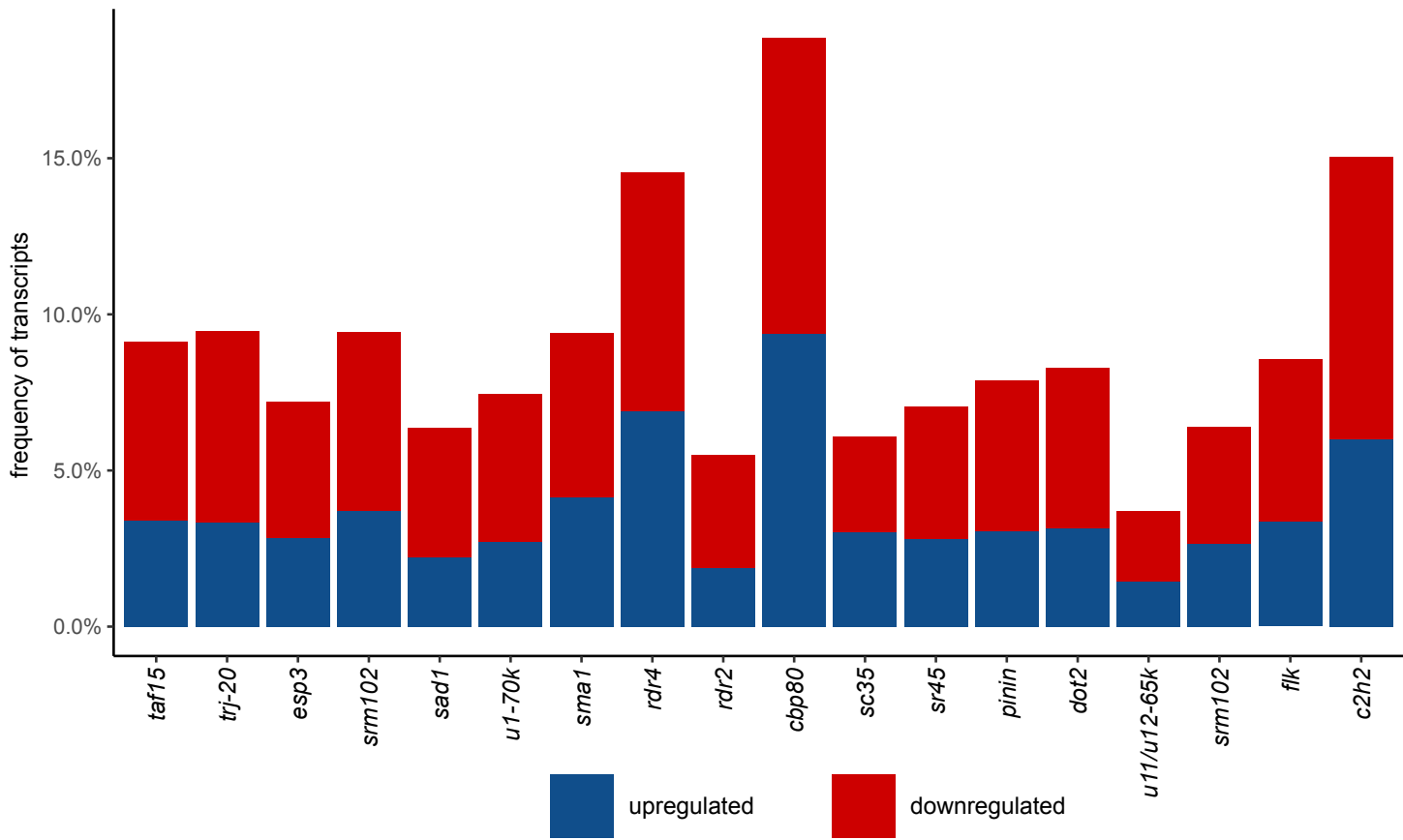
Level of gene expression in the wild-type plant and *cbp80*, *c2h2*, *flk* mutants (p-value ≤ 0.05 *, ≤ 0.01 **, ≤ 0.001 ***).



Supplementary Figure 8.

Accumulation level of circRNAs' linear counterparts in the wild-type and mutant plants

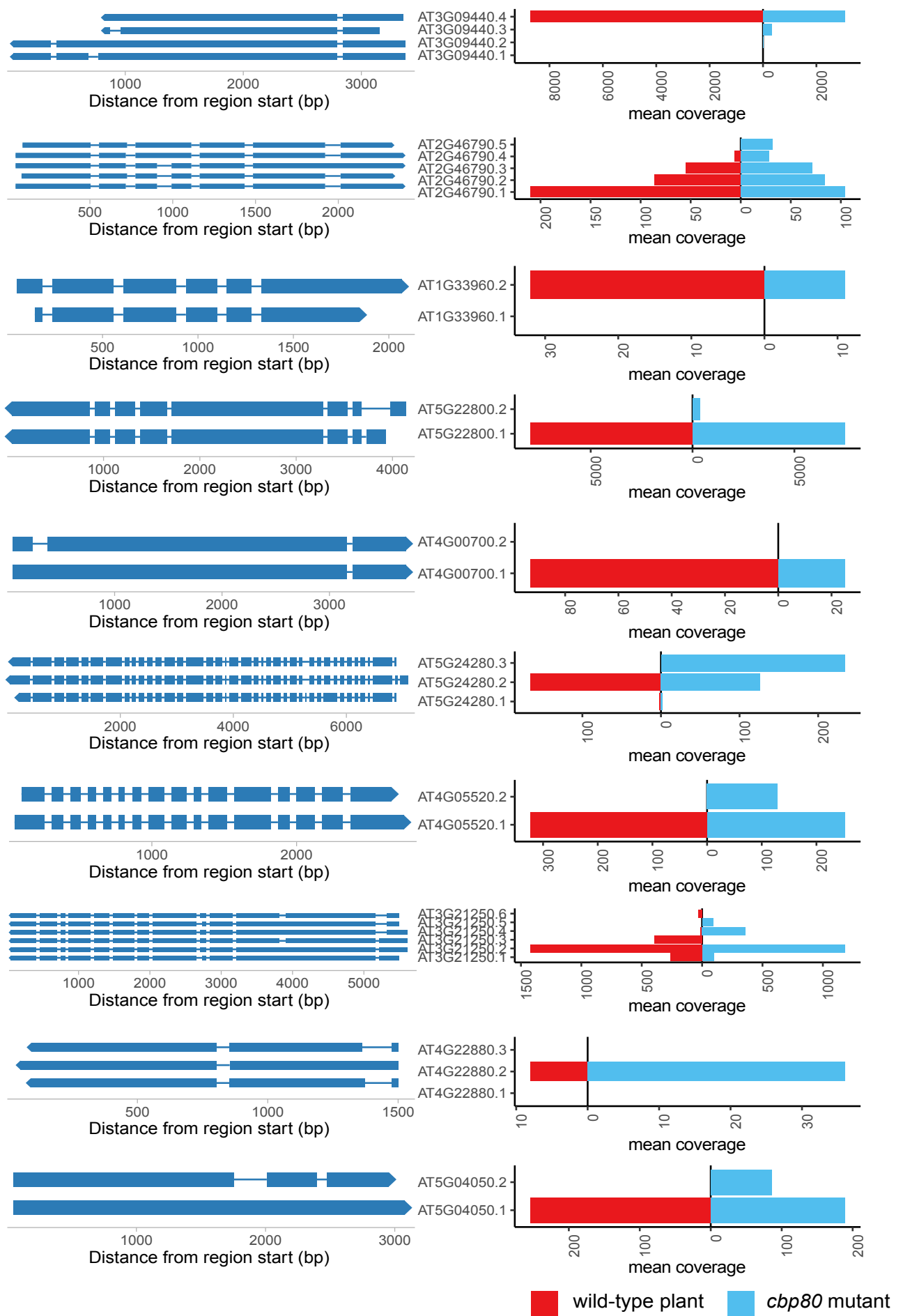
(p-value ≤ 0.05 *, ≤ 0.01 **, ≤ 0.001 ***).



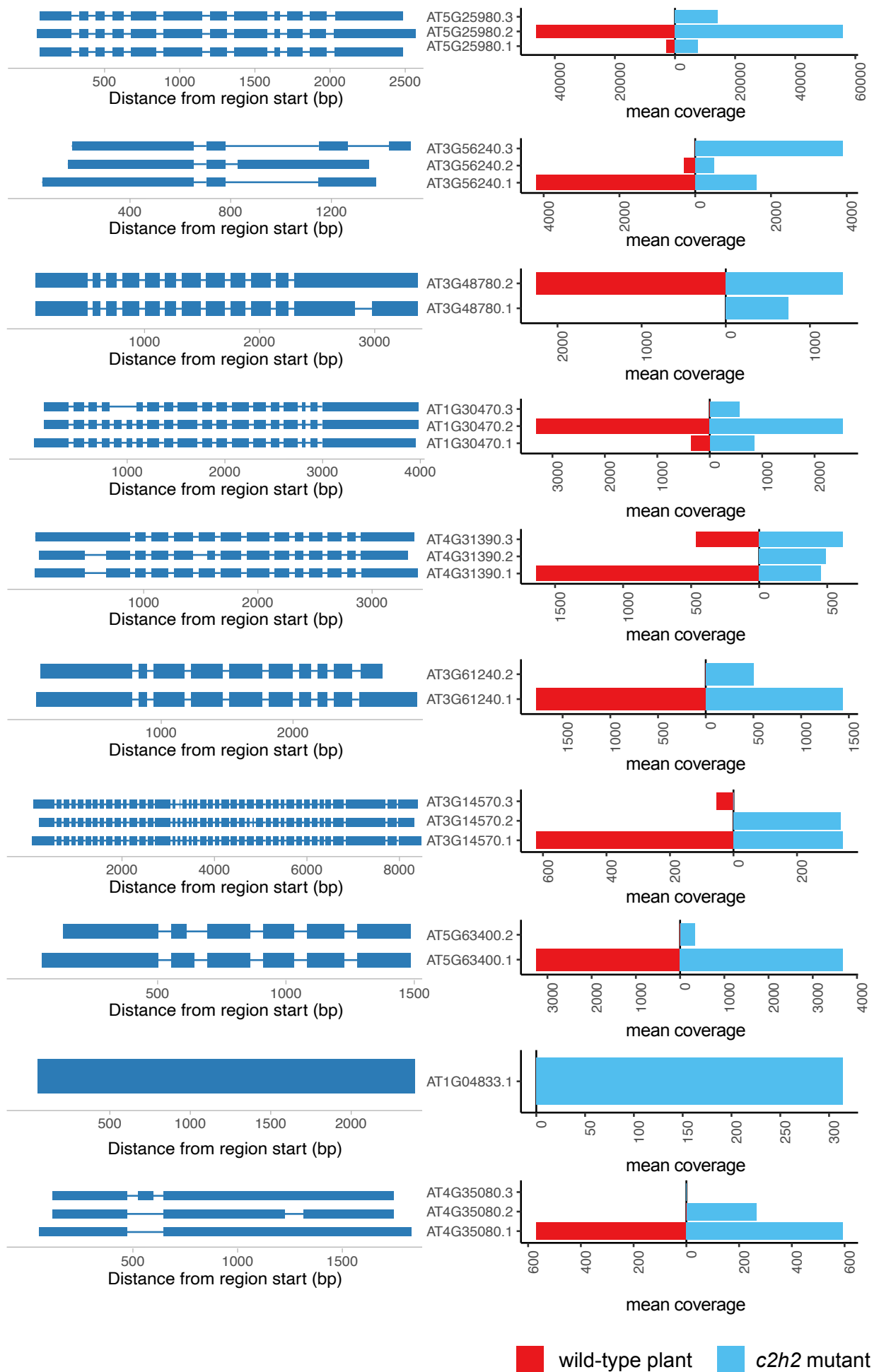
Supplementary Figure 9.

The percentage of differentially expressed transcripts in mutants compared to the wild-type plant.

(A)

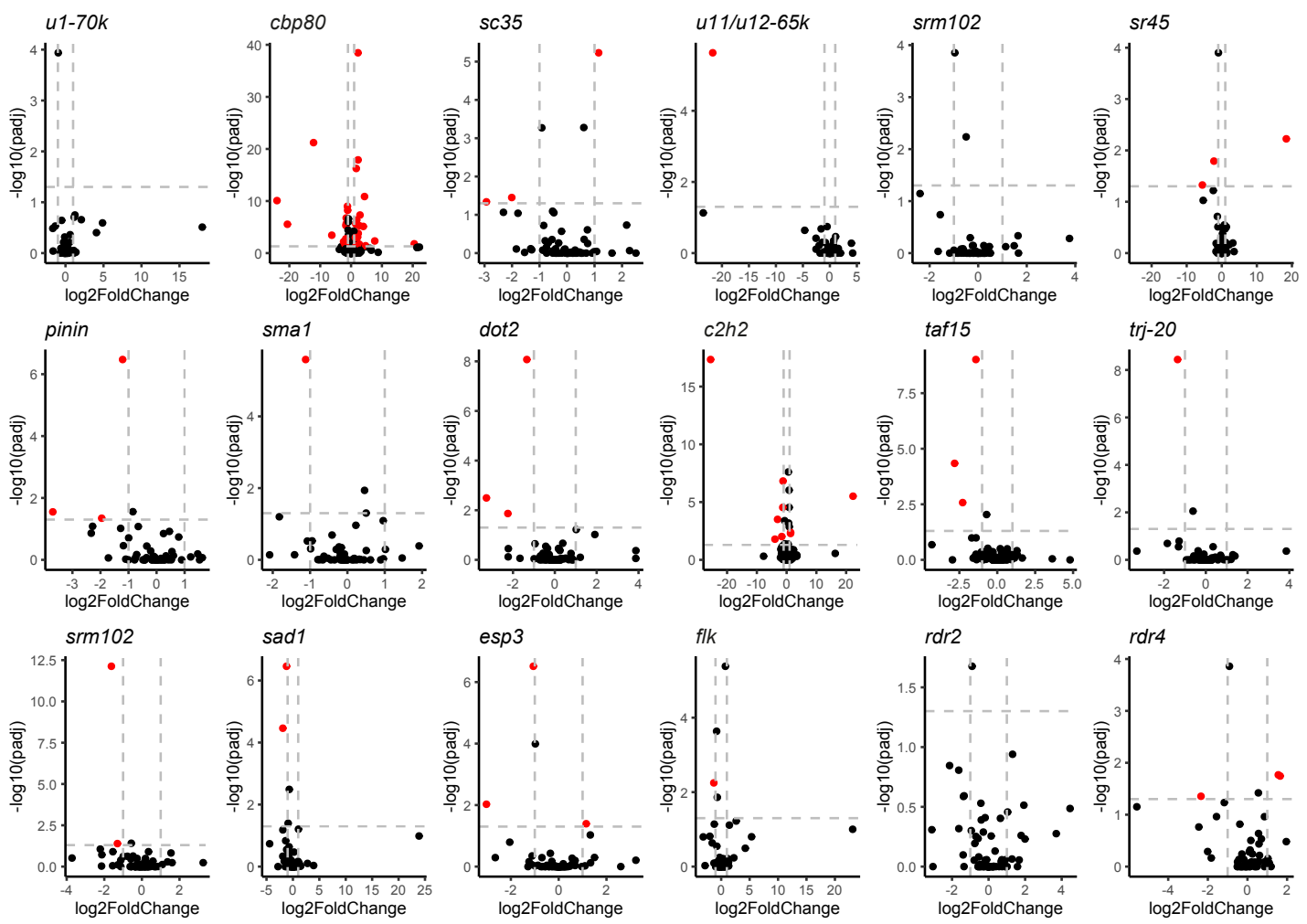


(B)



Supplementary Figure 10.

Top 10 differentially expressed transcripts for (A) *cbp80* and (B) *c2h2* mutants and their gene structure.



Supplementary Figure 11.

The expression profile of transcripts from genes generating circRNAs (identified in 4 biological replicates) in mutants vs. wild-type plants.

3. **Nowis K**, Jackowiak P, Figlerowicz M, Philips A;

At-C-RNA database, a one-stop source for information on circRNAs in Arabidopsis thaliana in a unified format

Database: The Journal of Biological Databases and Curation (2021),
<https://doi.org/10.1093/database/baab074>

Pięcioletni IF 4,8, MEiN 100

At-C-RNA database, a one-stop source for information on circRNAs in *Arabidopsis thaliana* in a unified format

Katarzyna Nowis, Paulina Jackowiak, Marek Figlerowicz and Anna Philips *

Institute of Bioorganic Chemistry, Polish Academy of Sciences, Z. Noskowskiego str. 12/14, Poznan 61-704, Poland

*Corresponding author: Tel: +48-61-852-85-03 (ext.1647); Email: aphilips@ibch.poznan.pl

Citation details: Nowis, K., Jackowiak, P., Figlerowicz, M. *et al.* At-C-RNA database, a one-stop source for information on circRNAs in *Arabidopsis thaliana* in a unified format. *Database* (2021) Vol. 2021: article ID baab074; DOI: <https://doi.org/10.1093/database/baab074>

Abstract

Circular RNAs (circRNAs) are a large class of noncoding RNAs with functions that, in most cases, remain unknown. Recent genome-wide analysis of circRNAs using RNA-Seq has revealed that circRNAs are abundant and some of them conserved in plants. Furthermore, it has been shown that the expression of circRNAs in plants is regulated in a tissue-specific manner. *Arabidopsis thaliana* circular RNA database is a new resource designed to integrate and standardize the data available for circRNAs in a model plant *A. thaliana*, which is currently the best-characterized plant in terms of circRNAs. The resource integrates all applicable publicly available RNA-seq datasets. These datasets were subjected to extensive reanalysis and curation, yielding results in a unified format. Moreover, all data were normalized according to our optimized approach developed for circRNA identification in plants. As a result, the database accommodates circRNAs identified across organs and seedlings of wild-type *A. thaliana* and its single-gene knockout mutants for genes related to splicing. The database provides free access to unified data and search functionalities, thus enabling comparative analyses of *A. thaliana* circRNAs between organs, variants and studies for the first time.

Database URL: <https://plantcircrna.ibch.poznan.pl/>

Introduction

Circular RNAs (circRNAs) are a class of noncoding alternatively spliced transcripts. It has been shown that circRNAs are present across the eukaryotic tree of life (1). Most efforts have been put into the identification and functional studies of circRNAs in animals (2) and humans (3–7). However, reliable identification and quantitation of plant circRNAs appear to be indispensable not only for the plant science field but also for the proper understanding of the universal rules that govern the formation and functioning of these RNAs across kingdoms and the significance of circRNAs in a broad evolutionary context.

The advent of RNA-Seq has driven the rapid expansion of circRNA studies. Next-generation sequencing technology can provide a comprehensive distribution of circRNAs in the whole organism and its particular organs. This situation is reflected in an increasing number of RNA-seq-based reports on plant circRNAs. Although *A. thaliana* circRNAs have been characterized in multiple studies, a comparison of their results reveals clear discrepancies. The main reason for this situation is a lack of standardization in the methods applied for circRNA analyses. For example, a large fraction of these molecules was identified based on RNA-seq data generated earlier to study gene expression levels. Moreover, the isolation, sequencing and bioinformatics protocols were rarely optimized for circRNA research and differed significantly between studies. This led to the situation that the

results published on circRNAs were inconsistent and impossible to comprehensively analyze. The data obtained thus far have been deposited in PlantCircBase (8), which encompasses 19 plant species, including *A. thaliana*, in PlantCircNet (9) and in AtCircDB (10), dedicated exclusively to circRNAs in *A. thaliana*. Unfortunately, circRNAs included in these databases come from different studies and were not curated, nor were their representations unified or normalized. None of these databases include the circRNAs identified in *A. thaliana* knockout mutants. Given the above, comprehensive comparative analyses of circRNAs in this model species have been significantly hampered. To change this situation, we reanalyzed RNA-seq raw data available in the public domain with our protocol for circRNA identification in plants (11) and developed At-C-RNA database to integrate and standardize the available circRNA data (see Figure 1).

Materials and methods

Data source

The At-C-RNA database consists of circRNAs identified by reanalyzing publicly available RNA-seq data. CircRNAs do not have polyA tail and thus can only be identified in datasets generated for rRNA-depleted libraries (and not polyA-selected). In search for relevant data, we browsed SRA NCBI with the following criteria: *A. thaliana* species,

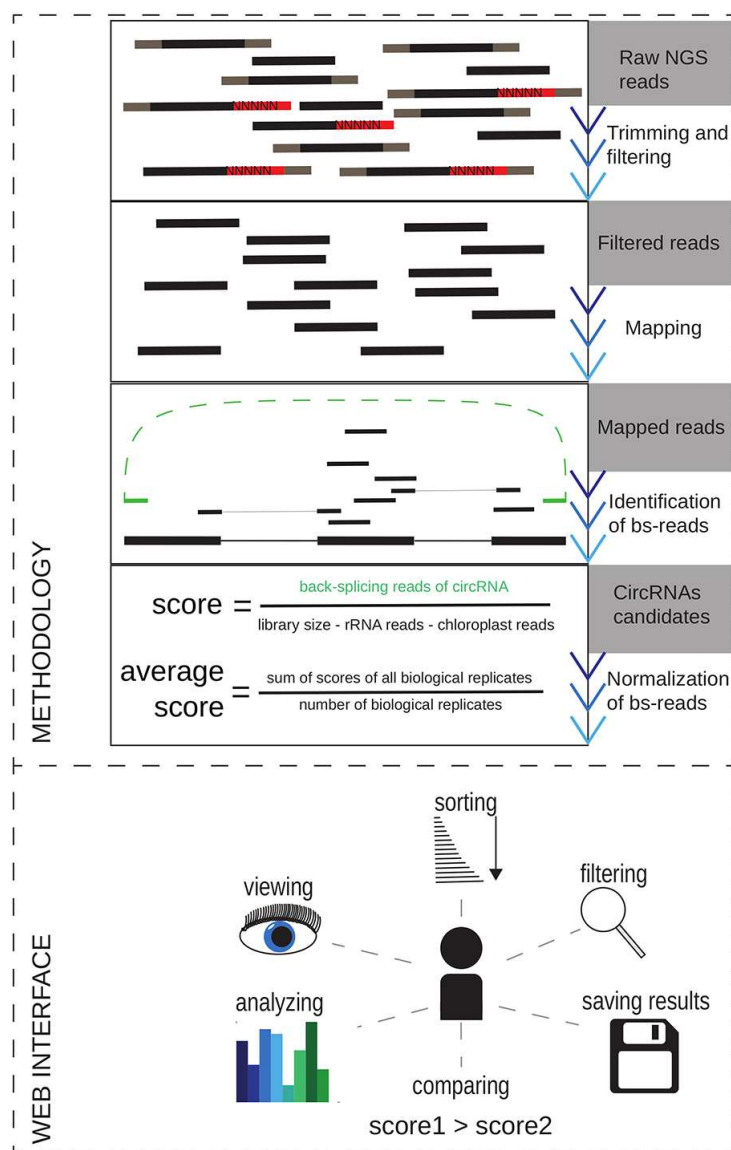


Figure 1. The overview of the methodology used for circRNAs reidentification in the publicly available datasets and the schematic view of the At-C-RNA utilities.

transcriptomic data, ncRNA, RNA-seq method, Illumina platform, paired-end library layout and rRNA-depleted data ('reduced representation' or 'inverse RNA selection', according to NCBI guidelines). SRA NCBI search query looked as follows: 'Arabidopsis thaliana'[Organism] AND 'transcriptomic'[Source] AND ('rna seq'[Strategy] OR 'ncrna seq'[Strategy]) AND 'platform illumina'[Properties] AND 'library layout paired'[Properties] AND ('reduced representation'[Selection] OR 'inverse rrna'[Selection]). Moreover, we compared the sources used by other plant circRNA databases and we chose those that met our criteria. In total, we utilized third-party data from eight studies (100 SRA files) and two datasets (110 SRA files) from our previous studies (11, 12). All analyzed datasets are presented in Table 1.

Web server implementation

The website was developed in an easy-to-use format with a responsive interface using bootstrap 4, jquery, and CSS technologies. The web framework was designed in Django (python 2.7.15). For table representation jsgrid-1.5.3, select2 was used. Charts showing data from the tables were created with the Google Charts tool and jvonn (13). Excel reports were generated using python packages xlswriter, pandas and NumPy.

Results

At-C-RNA content

Currently, in At-C-RNA, 113 327 circRNAs are deposited. Notably, only 19.7%, 18.9% and 16.2%, are reported

Table 1. The publicly available datasets utilized by At-C-RNA

BioProject	SRA IDs	Study
PRJNA525820	SRR11279578, SRR11279579, SRR11279580, SRR11279581, SRR11279582, SRR11279583, SRR11279584, SRR11279585, SRR11279586, SRR11279587, SRR11279588, SRR11279589, SRR11279590, SRR11279591, SRR11279592, SRR11279593, SRR11279594, SRR11279595, SRR11279596, SRR11279597, SRR11279598, SRR11279599, SRR11279600, SRR11279601, SRR11279602, SRR11279603, SRR11279604, SRR11279605, SRR11279606, SRR11279607, SRR11279608, SRR11279609, SRR11279610, SRR11279611	(11)
PRJNA630951	SRR11784202, SRR11784203, SRR11784204, SRR11784205, SRR11784206, SRR11784207, SRR11784208, SRR11784209, SRR11784210, SRR11784211, SRR11784212, SRR11784213, SRR11784214, SRR11784215, SRR11784216, SRR11784217, SRR11784218, SRR11784219, SRR11784220, SRR11784221, SRR11784222, SRR11784223, SRR11784224, SRR11784225, SRR11784226, SRR11784227, SRR11784228, SRR11784229, SRR11784230, SRR11784231, SRR11784232, SRR11784233, SRR11784234, SRR11784235, SRR11784236, SRR11784237, SRR11784238, SRR11784239, SRR11784240, SRR11784241, SRR11784242, SRR11784243, SRR11784244, SRR11784245, SRR11784246, SRR11784247, SRR11784248, SRR11784249, SRR11784250, SRR11784251, SRR11784252, SRR11784253, SRR11784254, SRR11784255, SRR11784256, SRR11784257, SRR11784258, SRR11784259, SRR11784260, SRR11784261, SRR11784262, SRR11784263, SRR11784264, SRR11784265, SRR11784266, SRR11784267, SRR11784268, SRR11784269, SRR11784270, SRR11784271, SRR11784272, SRR11784273, SRR11784274, SRR11784275, SRR11784276, SRR11784277	(12)
PRJDB6099	DRR099080, DRR099081, DRR099082, DRR099083, DRR099084, DRR099085	NA
PRJNA218215	SRR1004790, SRR1004791, SRR1004829, SRR1004830, SRR1004831, SRR1004832, SRR1004833, SRR1004834	(14)
PRJNA596364	SRR10727148, SRR10727149, SRR10727150, SRR10727151, SRR10727152, SRR10727153, SRR10727154, SRR10727155, SRR10727156	(15)
PRJNA186843	SRR2079771, SRR2079772, SRR2079773, SRR2079774, SRR2079775, SRR2079776, SRR2079777, SRR2079778, SRR2079779, SRR2079780, SRR2079781, SRR2079782, SRR2079783, SRR2079784, SRR2079785, SRR2079786, SRR2079787, SRR2079788, SRR2079789, SRR2079790, SRR2079791, SRR2079792, SRR2079793, SRR2079794, SRR2079795, SRR2079796, SRR2079797, SRR2079798, SRR2079799, SRR2079800, SRR2079801, SRR2079802, SRR2079803, SRR2079804, SRR2079805, SRR2079806, SRR2079807, SRR2079808, SRR2079809, SRR2079810, SRR2079811, SRR2079812, SRR2079813, SRR2079814, SRR2079815, SRR2079816, SRR2079817, SRR2079818, SRR2079819, SRR2079820, SRR2079821, SRR2079822, SRR2079823, SRR2079824, SRR2079825, SRR2079826, SRR2079827, SRR2079828, SRR2079829, SRR2079830, SRR2079831, SRR2079832, SRR2079833, SRR2079834, SRR2079835, SRR2079836, SRR2079837, SRR2079838, SRR2079839, SRR2079840, SRR2079841, SRR2079842, SRR2079843, SRR2079844, SRR2079845, SRR2079846, SRR2079847, SRR2079848, SRR2079849, SRR2079850, SRR2079851, SRR2079852, SRR2079853, SRR2079854, SRR2079855, SRR2079856, SRR2079857, SRR2079858, SRR2079859, SRR2079860, SRR2079861, SRR2079862, SRR2079863, SRR2079864, SRR2079865, SRR2079866, SRR2079867, SRR2079868, SRR2079869, SRR2079870, SRR2079871, SRR2079872, SRR2079873, SRR2079874, SRR2079875, SRR2079876, SRR2079877, SRR2079878, SRR2079879, SRR2079880, SRR2079881, SRR2079882, SRR2079883, SRR2079884, SRR2079885, SRR2079886, SRR2079887, SRR2079888, SRR2079889, SRR2079890, SRR2079891, SRR2079892, SRR2079893, SRR2079894, SRR2079895, SRR2079896, SRR2079897, SRR2079898, SRR2079899, SRR2079900, SRR2079901, SRR2079902, SRR2079903, SRR2079904, SRR2079905, SRR2079906, SRR2079907, SRR2079908, SRR2079909, SRR2079910, SRR2079911, SRR2079912, SRR2079913, SRR2079914, SRR2079915, SRR2079916, SRR2079917, SRR2079918, SRR2079919, SRR2079920, SRR2079921, SRR2079922, SRR2079923, SRR2079924, SRR2079925, SRR2079926, SRR2079927, SRR2079928, SRR2079929, SRR2079930, SRR2079931, SRR2079932, SRR2079933, SRR2079934, SRR2079935, SRR2079936, SRR2079937, SRR2079938, SRR2079939, SRR2079940, SRR2079941, SRR2079942, SRR2079943, SRR2079944, SRR2079945, SRR2079946, SRR2079947, SRR2079948, SRR2079949, SRR2079950, SRR2079951, SRR2079952, SRR2079953, SRR2079954, SRR2079955, SRR2079956, SRR2079957, SRR2079958, SRR2079959, SRR2079960, SRR2079961, SRR2079962, SRR2079963, SRR2079964, SRR2079965, SRR2079966, SRR2079967, SRR2079968, SRR2079969, SRR2079970, SRR2079971, SRR2079972, SRR2079973, SRR2079974, SRR2079975, SRR2079976, SRR2079977, SRR2079978, SRR2079979, SRR2079980, SRR2079981, SRR2079982, SRR2079983, SRR2079984, SRR2079985, SRR2079986, SRR2079987, SRR2079988, SRR2079989, SRR2079990, SRR2079991, SRR2079992, SRR2079993, SRR2079994, SRR2079995, SRR2079996, SRR2079997, SRR2079998, SRR2079999, SRR2080000	(16)
PRJNA311178	SRR3151787, SRR5591021, SRR5591022, SRR5591023, SRR5591024, SRR5591025, SRR5591026, SRR5591027, SRR5591028, SRR5591029, SRR5591030, SRR5591031, SRR5591032, SRR5591033	(17)
PRJNA437291	SRR6814509	(18)
PRJNA511671	SRR8368644, SRR8368646, SRR8368647, SRR8368649, SRR8368650, SRR8368652, SRR8368653, SRR8368635, SRR8368636, SRR8368637, SRR8368638, SRR8368639, SRR8368640, SRR8368641, SRR8368642, SRR8368643, SRR8368645, SRR8368648, SRR8368651, SRR8368654	(19)
PRJEB32782	ERR3489926, ERR3489927, ERR3489928, ERR3489929, ERR3489930, ERR3489931, ERR3489932, ERR3489933, ERR3489934, ERR3489935, ERR3489936, ERR3489937, ERR3489938, ERR3489939	(20)

in PlantcircBase (8), PlantcircNet (9) and AtCircDB (10), respectively (access: 2 September 2021). Each circRNA was assigned a unique identifier according to the common pattern AT_chromosome_number:circRNA_start-circRNA_stop, making comparisons between studies/other databases possible and convenient. The following information is available for each circRNA: (i) the study in which the datasets were generated, (ii) plant line and organ in which circRNAs were identified, (iii) the average score, computed for all circRNAs according to the same procedure, (iv) the individual component scores and (v) information regarding whether a circRNA was confirmed with RNase R experiments. It is worth mentioning that we introduced a novel ‘reproducibility’ criterion in At-C-RNA, as in our previous studies, and we showed that most circRNAs in *A. thaliana* are produced spontaneously and thus possibly carry no biological function (11). It is important to highlight the reproducible circRNAs, which may have functional potential and thus are especially important to the wide range of studies. We defined that a circRNA is reproducible if it was identified in at least four biological replicates in a specific organ and line within a study.

In total, 655 circRNAs were classified as reproducible (see Figure 2A). Of these, 226 were identified in all of the analyzed plant tissues (flower, leaf, root and seedling) and the whole plant. The highest number of unique reproducible circRNAs (35) was found in the leaf. On the contrary, two other organs, root and flower, revealed only two and one circRNAs typical only to this tissue, respectively. No unique circRNA

was found in the seedling. Most of the genes (362) giving rise to the reproducible circRNAs produced above five circRNAs isoforms and only 11 genes produced one circRNA isoform (see Figure 2B). Most reproducible circRNAs (91.6%) score ranges from 1 to 15 what corresponds to the rather low abundance, which most circRNAs display. Fifty-five of circRNAs (8.4%) exceeded an average score over 15 (see Figure 2C). The distribution of reproducible circRNAs on the chromosomes is shown in Figure 2D. Most reproducible circRNAs originated from genes located on chromosome 1 and none from mitochondrial.

At-C-RNA database utility

Novelty

At-C-RNA is the only resource where data across different studies have been reanalyzed, standardized and unified. Moreover, At-C-RNA is the only database that provides information on circRNA reproducibility and occurrence in both wild-type and mutant *A. thaliana* plants.

Data browsing and filtering

The At-C-RNA database aggregates circRNAs in a table that allows the user to define the filtering criteria. Data sorting by each column is possible. Each column has a window where filtering criteria can be typed. There is also a possibility to manually delete selected circRNAs from the table. Moreover,

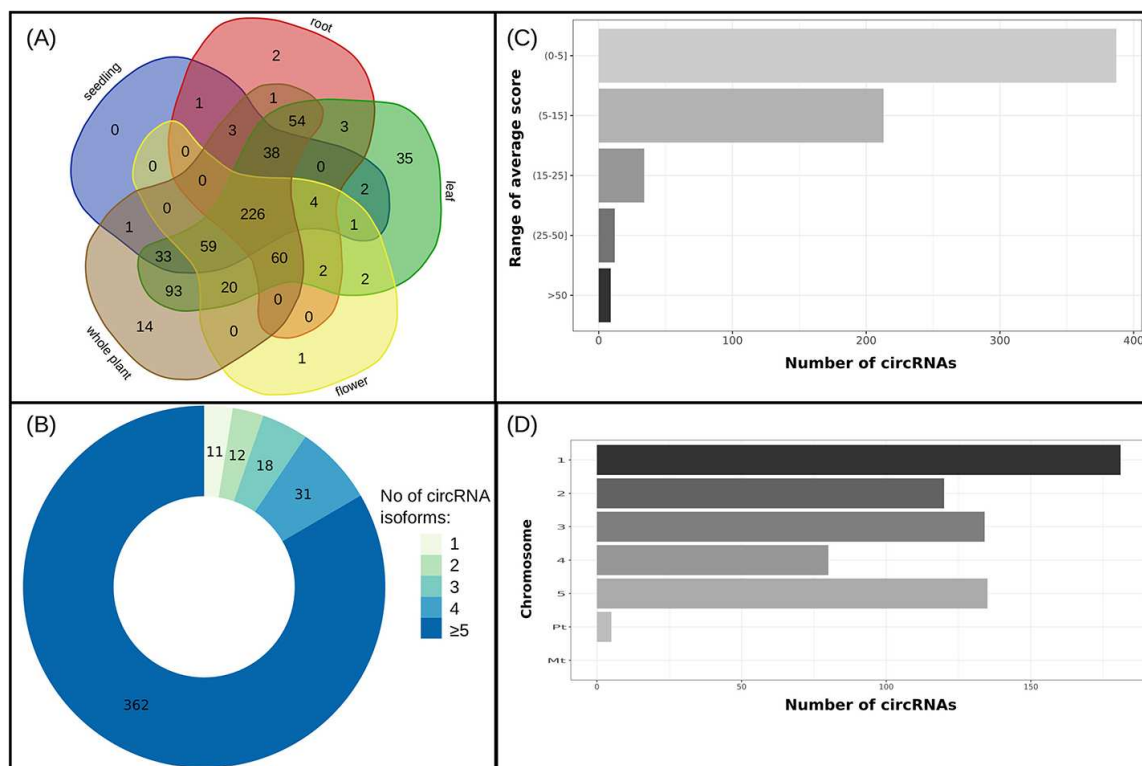


Figure 2. (A) Distribution of reproducible circRNAs across organs/seedling and the whole plant, (B) circRNAs' isoform number of genes producing reproducible circRNAs, (C) average score ranges for reproducible circRNAs and (D) distribution of reproducible circRNAs across chromosomes.

users can also filter the table by clicking on the interactive charts below the table.

Data download

A previously filtered collection of circRNAs can be downloaded with the 'Excel Report' button. Moreover, users can create pivot tables and generate plots from columns of interest.

Genomic region and gene information

The database also holds information from external databases (i.e. Ensembl and NCBI). CircRNA ID redirects the user to the Ensembl genome browser where the genomic region of the circRNA of interest can be explored. Moreover, users can read extended information about genes using an external link to NCBI.

Common use cases

A frequent task is to search for circRNAs that are produced in a reproducible manner, as only such molecules may carry biological functions. At-C-RNA is the only database where data were curated, and a reproducibility measure was defined for each circRNA. A default filter on the table enables the visualization of these reproducible circRNAs. Moreover, our database enables multilevel filtering, for example, users can filter the data table showing only reproducible circRNAs from a specific gene that are confirmed in RNase R-treated samples.

Data curation

All circRNAs deposited in At-C-RNA were reidentified from raw data by our in-home protocol developed for circRNAs identification in plants. We plan to update the database and successively reanalyze and add new circRNAs data, as they appear in the public domain.

Discussion

Currently, At-C-RNA is the biggest resource of circRNAs in *A. thaliana*, encompassing 113 327 circRNAs. The At-C-RNA database provides not only a comprehensive and convenient source of unified information on the circRNAs in *A. thaliana* but also a user-friendly interface that allows the user to run analyses, the results of which are available in the form of interactive graphical reports and summaries. This platform can be used in plant circRNA research as well as in all studies that focus on the general features of circRNAs and explore the functional potential of these molecules. By unifying data and providing essential tools, At-C-RNA is a robust platform for comparative analyses of circRNAs. The resource will be curated—we plan to successively reanalyze and add new data from the public domain. We believe that At-C-RNA resources will not only contribute to studies on circRNAs biogenesis and function in plants but also will help to understand the universal rules that govern the formation and functioning of circRNAs and their significance in a broad evolutionary context.

Acknowledgements

The authors thank Michal Stelmaszczuk for making photos placed in At-C-RNA and Marcin Osuch for designing the database logo.

Funding

Polish National Science Centre grant (UMO-2014/15/D/NZ2/02305) to A.P.

Conflict of interest

The authors declare that they have no competing interests.

Data availability

Freely available at <https://plantcircrna.ibch.poznan.pl/>. Website implemented in Django, MySQL and Apache, with responsive design and all major browsers supported.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Authors' contributions

A.P. conceived the overall idea of the study. A.P., P.J. and M.F. designed the study and discussed the results. A.P. and K.N. performed bioinformatics analyses. K.N. built the web application. K.N., A.P. and P.J. drafted the manuscript. A.P. was responsible for the final version of the manuscript. All authors contributed to the article and approved the submitted version.

References

1. Wang,P.L., Bao,Y., Yee,M.C. *et al.* (2014) Circular RNA is expressed across the eukaryotic tree of life. *PLoS One*, **9**, e90859.
2. Memczak,S., Jens,M., Elefsinioti,A. *et al.* (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, **495**, 333–338.
3. Czubak,K., Taylor,K., Piasecka,A. *et al.* (2019) Global increase in circular RNA levels in myotonic dystrophy. *Front Genet.*, **10**, 649.
4. Liu,J., Li,D., Luo,H. *et al.* (2019) Circular RNAs: the star molecules in cancer. *Mol. Aspects Med.*, **70**, 141–152.
5. Qu,S., Liu,Z., Yang,X. *et al.* (2018) The emerging functions and roles of circular RNAs in cancer. *Cancer Lett.*, **414**, 301–309.
6. Hansen,T.B., Jensen,T.I., Clausen,B.H. *et al.* (2013) Natural RNA circles function as efficient microRNA sponges. *Nature*, **495**, 384–388.
7. Peng,Y., Song,X., Zheng,Y. *et al.* (2018) circCOL3A1-859267 regulates type I collagen expression by sponging miR-29c in human dermal fibroblasts. *Eur. J. Dermatol.*, **28**, 613–620.
8. Chu,Q., Zhang,X., Zhu,X. *et al.* (2017) PlantcircBase: a database for plant circular RNAs. *Mol Plant*, **10**, 1126–1128.
9. Zhang,P., Meng,X., Chen,H. *et al.* (2017) PlantCircNet: a database for plant circRNA-miRNA-mRNA regulatory networks. *Database (Oxford)*, **2017**, bax089.
10. Ye,J., Wang,L., Li,S. *et al.* (2017) AtCircDB: a tissue-specific database for Arabidopsis circular RNAs. *Brief Bioinform.*, **20**, 58–65.
11. Philips,A., Nowis,K., Stelmaszczuk,M. *et al.* (2020) Expression landscape of circRNAs in Arabidopsis thaliana seedlings and adult tissues. *Front Plant Sci.*, **11**, 576581.
12. Philips,A., Nowis,K., Stelmaszczuk,M. *et al.* (2020) Arabidopsis thaliana cbp80, c2h2, and flk knockout mutants accumulate increased amounts of circular RNAs. *Cells*, **9**, 1937.
13. Bardou,P., Mariette,J., Escudie,F. *et al.* (2014) jvenn: an interactive Venn diagram viewer. *BMC Bioinform.*, **15**, 293.
14. Ye,C.Y., Chen,L., Liu,C. *et al.* (2015) Widespread noncoding circular RNAs in plants. *New Phytol.*, **208**, 88–95.
15. Sun,Z., Huang,K., Han,Z. *et al.* (2020) Genome-wide identification of Arabidopsis long noncoding RNAs in response to the blue light. *Sci. Rep.*, **10**, 6229.
16. Liu,T., Zhang,L., Chen,G. *et al.* (2017) Identifying and characterizing the circular RNAs during the lifespan of Arabidopsis leaves. *Front Plant Sci.*, **8**, 1278.
17. Chen,G., Cui,J., Wang,L. *et al.* (2017) Genome-wide identification of circular RNAs in Arabidopsis thaliana. *Front Plant Sci.*, **8**, 1678.
18. Coate,J.E., Schreyer,W.M., Kum,D. *et al.* (2020) Robust cytonuclear coordination of transcription in nascent Arabidopsis thaliana autopolyploids. *Genes (Basel)*, **11**, 134.
19. Zhang,P., Fan,Y., Sun,X. *et al.* (2019) A large-scale circular RNA profiling reveals universal molecular mechanisms responsive to drought stress in maize and Arabidopsis. *Plant J.*, **98**, 697–713.
20. Parker,M.T., Knop,K., Sherwood,A.V. *et al.* (2020) Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m(6)A modification. *Elife*, **9**, e49658.

OŚWIADCZENIA DOTYCZĄCE PRAC WCHODZĄCYCH W
SKŁAD ROZPRAWY DOKTORSKIEJ

prof. dr hab. Marek Figlerowicz

Poznań, 17.12.2021 r.

OŚWIADCZENIE

Dotyczy rozprawy doktorskiej mgr Katarzyny Nowis

Mgr Katarzyna Nowis wykonywała pracę doktorską w Instytucie Chemii Bioorganicznej PAN. Poniżej przedstawiam zakres prac wykonywanych przez mgr Katarzynę Nowis w publikacjach, których jestem autorem do korespondencji:

Oświadczam, że:

1. wkład mgr Katarzyny Nowis w powstanie przedstawionej poniżej pracy polegał na: wykonaniu bioinformatycznych analiz, w szczególności: identyfikacji kolistych RNA (circRNA) na podstawie danych z RNA-seq w 3 organach (kwiecie, liściu i korzeniu) oraz siewce u *Arabidopsis thaliana*, określeniu dystrybucji sekwencji kodujących circRNA w chromosomach, określeniu zawartości rRNA w danych z RNA-seq, porównaniu liczby powtarzalnych i unikatowych circRNA w 2 typach bibliotek RNA-seq i 4 powtórzeniach biologicznych, walidacja 6 metod normalizacji danych RNA-seq, porównaniu akumulacji circRNA w 3 organach i siewkach, analizie korelacji pomiędzy akumulacją circRNA i ich liniowych odpowiedników, stworzeniu rysunków dotyczących bioinformatycznej części pracy oraz udziale w planowaniu eksperymentów oraz przygotowywaniu manuskryptu.

Philips A, **Nowis K**, Stelmaszczuk M, Jackowiak P, Podkowiński J, Handschuh L, Figlerowicz M; Expression Landscape of circRNAs in Arabidopsis thaliana Seedlings and Adult Tissues Frontiers in Plant Science (2020), <https://doi.org/10.3389/fpls.2020.576581>

2. wkład mgr Katarzyny Nowis w powstanie przedstawionej poniżej pracy polegał na: wykonaniu analiz bioinformatycznych, w szczególności identyfikacji circRNA na podstawie danych z RNA-seq w 18 mutantach i typie dzikim u *A. thaliana*, identyfikacji puli unikatowych circRNA specyficznych dla danego mutantu, normalizacji danych z RNA-seq, określeniu poziomu akumulacji circRNA w mutantach i typie dzikim, porównaniu akumulacji circRNA pomiędzy mutantami i typem dzikim, określeniu dystrybucji egzonów od których zaczyna się produkcja circRNA w poszczególnych mutantach i typie dzikim, określeniu poziomu ekspresji genów, stworzeniu rysunków oraz udziale w planowaniu eksperymentów oraz przygotowywaniu manuskryptu.

Philips A, **Nowis K**, Stelmaszczuk M, Podkowiński J, Handschuh L, Jackowiak P, Figlerowicz M; Arabidopsis thaliana cbp80, c2h2, and flk Knockout Mutants Accumulate Increased Amounts of Circular RNAs Cells (2020), <https://doi.org/10.3390/cells9091937>



dr hab. Anna Philips

Poznań, 19.08.2022 r.

OŚWIADCZENIE

Dotyczy rozprawy doktorskiej mgr Katarzyny Nowis

Mgr Katarzyna Nowis realizowała pracę doktorską w Instytucie Chemii Bioorganicznej PAN. Poniżej przedstawiam zakres prac wykonanych przez mgr Katarzynę Nowis w publikacjach, które wchodzi w skład jej rozprawy doktorskiej.

Oświadczam, że w pracy:

1. Philips A, **Nowis K**, Stelmaszczuk M, Jackowiak P, Podkowiński J, Handschuh L, Figlerowicz M; Expression Landscape of circRNAs in Arabidopsis thaliana Seedlings and Adult Tissues *Frontiers in Plant Science* (2020), <https://doi.org/10.3389/fpls.2020.576581>

wkład mgr Katarzyny Nowis polegał na wykonaniu analiz bioinformatycznych, tj. identyfikacji, analizie jakościowej i ilościowej kolistych RNA na podstawie danych RNA-seq. Określeniu pul circRNA w różnych tkankach rośliny i ich analiza porównawcza. Walidacji sześciu metod normalizacji danych RNA-seq pod kątem analiz circRNA. Stworzeniu (rysunki 2-6 i suplementarny rysunek 5) oraz tabel, a także udziale w planowaniu eksperymentów i przygotowaniu manuskryptu.

2. Philips A, **Nowis K**, Stelmaszczuk M, Podkowiński J, Handschuh L, Jackowiak P, Figlerowicz M; Arabidopsis thaliana cbp80, c2h2, and flk Knockout Mutants Accumulate Increased Amounts of Circular RNAs *Cells* (2020), <https://doi.org/10.3390/cells9091937>

wkład mgr Katarzyny Nowis polegał na wykonaniu analiz bioinformatycznych, tj. identyfikacji i analizie circRNA na podstawie danych RNA-seq w 18 mutantach i typie dzikim *A. thaliana*. Określeniu budowy circRNA. Określeniu poziomu ekspresji genów. Stworzeniu rysunków i tabel, a także udziale w planowaniu eksperymentów i przygotowaniu manuskryptu.

3. **Nowis K**, Jackowiak P, Figlerowicz M, Philips A; At-C-RNA database, a one-stop source for information on circRNAs in Arabidopsis thaliana in a unified format *Database: The Journal of Biological Databases and Curation* (2021), <https://doi.org/10.1093/database/baab074>

wkład mgr Katarzyny Nowis polegał na wykonaniu analiz bioinformatycznych, w szczególności analizie jakościowej i ilościowej circRNA, stworzeniu rysunków, zaprojektowaniu i stworzeniu bazy danych i aplikacji internetowej oraz udziale w planowaniu prac oraz przygotowaniu manuskryptu.

dr hab. Anna Philips



PODPIS ZAUFANY

ANNA
PHILIPS
01.09.2022 13:35:11 [GMT+2]
Dokument podpisany elektronicznie
podpisem zaufanym

Katarzyna Nowis

Poznań, 19.08.2022 r.

OŚWIADCZENIE

Poniżej przedstawiam zakres prac wykonywanych przeze mnie w publikacjach, które wchodzą w skład mojej rozprawy doktorskiej oraz oświadczam, że:

1. W poniższej pracy mój wkład polegał na wykonaniu analiz bioinformatycznych na podstawie danych RNA-seq, tj.:
 - 1.1. zidentyfikowałam koliste RNA (circRNA) w trzech organach (kwiatach, liściach, korzeniach) oraz siewkach *Arabidopsis thaliana*
 - 1.2. określiłam rozkład circRNA w chromosomach i określiłam ilość rRNA w próbkach
 - 1.3. porównałam powtarzalne i unikatowe circRNA w dwóch typach bibliotek i czterech powtórzeniach biologicznych
 - 1.4. zwalidowałam sześć metod normalizacji danych RNA-seq pod kątem analiz circRNA w oparciu o PCR emulsyjny (ddPCR)
 - 1.5. porównałam akumulację circRNA w trzech organach i siewkach
 - 1.6. zanalizowałam korelację pomiędzy akumulacją circRNA a akumulacją ich liniowych odpowiedników

Stworzyłam rysunki dotyczące bioinformatycznej części pracy (rysunki 2-6 i suplementarny rysunek 5) oraz tabele, a także brałam czynny udział w planowaniu eksperymentów i pisaniu manuskryptu.

Philips A, **Nowis K**, Stelmaszczuk M, Jackowiak P, Podkowiński J, Handschuh L, Figlerowicz M;
Expression Landscape of circRNAs in Arabidopsis thaliana Seedlings and Adult Tissues
Frontiers in Plant Science (2020), <https://doi.org/10.3389/fpls.2020.576581>

2. W poniższej pracy mój wkład polegał na wykonaniu analiz bioinformatycznych na podstawie danych RNA-seq tj.:
 - 2.1. zidentyfikowałam circRNA w 18 mutantach i typie dzikim *A. thaliana*
 - 2.2. zidentyfikowałam pulę unikatowych circRNA specyficznych dla danego mutantu
 - 2.3. znormalizowałam dane RNA-seq i określiłam poziom akumulacji circRNA w mutantach i typie dzikim
 - 2.4. porównałam akumulację circRNA pomiędzy mutantami i typem dzikim
 - 2.5. określiłam od których egzonów tworzone są circRNA w poszczególnych mutantach i typie dzikim
 - 2.6. określiłam poziom ekspresji genów w mutantach i typie dzikimStworzyłam rysunki i tabele, a także brałam udział w planowaniu eksperymentów i pisaniu manuskryptu.

Nowis

Philips A, **Nowis K**, Stelmaszczuk M, Podkowiński J, Handschuh L, Jackowiak P, Figlerowicz M;
Arabidopsis thaliana cbp80, c2h2, and flk Knockout Mutants Accumulate Increased Amounts of Circular RNAs
Cells (2020), <https://doi.org/10.3390/cells9091937>

3. W poniższej pracy mój wkład polegał na wykonaniu analiz bioinformatycznych (analizy jakościowej i ilościowej circRNA). Stworzyłam rysunki, zaprojektowałam i stworzyłam bazę danych i aplikację internetową. Brałam także udział w planowaniu prac oraz przygotowaniu i pisaniu manuskryptu.

Nowis K, Jackowiak P, Figlerowicz M, Philips A;
At-C-RNA database, a one-stop source for information on circRNAs in Arabidopsis thaliana in a unified format
Database: The Journal of Biological Databases and Curation (2021),
<https://doi.org/10.1093/database/baab074>

Katarzyna Nowis