

遺伝子制御ネットワーク推定を用いた
Pv11 細胞における
乾燥耐性制御機構に関する研究

2020 年度

山田 貴大

学位論文 博士（理学）

遺伝子制御ネットワーク推定を用いた
Pv11 細胞における
乾燥耐性制御機構に関する研究

2020 年度

慶應義塾大学大学院理工学研究科

山田 貴大

目次

第1章 緒言	1
1.1 本論文の構成	1
1.2 乾燥耐性	2
1.2.1 乾燥耐性獲得期	6
1.2.1.1 代謝	6
1.2.1.2 シグナル伝達	9
1.2.1.3 遺伝子制御	9
1.2.1.4 乾燥耐性獲得期のまとめ	13
1.2.2 無代謝期	15
1.2.3 再水和期	15
1.2.3.1 代謝	15
1.2.3.2 シグナル伝達	16
1.2.3.3 遺伝子制御	16
1.2.3.4 再水和期のまとめ	16
1.2.4 乾燥耐性に関するまとめ	16
1.3 遺伝子制御ネットワーク	18
1.3.1 遺伝子制御ネットワークの定義と研究事例	18
1.3.2 転写制御ネットワーク推定	25
1.3.2.1 構造学習型転写制御ネットワーク推定	25
1.3.2.2 回帰型転写制御ネットワーク推定	28
1.3.3 末端遺伝子制御ネットワーク推定	32
1.3.3.1 ChIP-seq データを用いた末端遺伝子制御ネットワーク推定	32
1.3.3.2 結合配列モチーフデータを用いた末端遺伝子制御ネットワーク推定	33
1.3.4 遺伝子制御ネットワーク推定に関するまとめ	35
1.4 本研究の目的	35
第2章 Pv11 細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の推定	37
2.1 背景	37
2.2 方法	38
2.2.1 CAGE-seq データの前処理	38

2.2.2	発現変動遺伝子解析	40
2.2.3	Heatmap および階層的クラスタリングによる時系列遺伝子発現パターンの可視化	40
2.2.4	ネムリユスリカゲノムに対する GO アノテーションおよび Fisher の正確確率検定に基づくエンリッチメント解析	40
2.2.5	タンパク質ドメイン構造解析	41
2.2.6	Pv11 細胞, ネムリユスリカ虫体および乾燥耐性を持たないヤモンユスリカの比較解析	41
2.2.7	RT qPCR	41
2.2.8	CAGE-seq データの可用性	42
2.3	結果	42
2.3.1	CAGE-seq データに基づく遺伝子発現量解析結果	42
2.3.2	発現変動遺伝子に対する GO 解析結果	46
2.3.3	乾燥耐性獲得期に対する解析結果	49
2.3.4	再水和期に対する解析結果	52
2.3.5	Pv11 細胞, ネムリユスリカおよびヤモンユスリカの発現変動遺伝子に関する比較解析結果	57
2.4	考察	60
2.4.1	別種との比較解析から考えられる Pv11 細胞とネムリユスリカの乾燥耐性の類似点及び相違点	60
2.4.2	本研究から得られた Pv11 細胞の乾燥耐性機構と構成する遺伝子群	63
第 3 章 Pv11 細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定		64
3.1	背景	64
3.2	方法	66
3.2.1	提供された RNA-seq データの概要及びデータ前処理	66
3.2.2	主成分分析	68
3.2.3	発現変動遺伝子解析	68
3.2.4	時系列発現に基づく転写制御ネットワーク推定	68
3.2.5	協調発現モジュールの検出と Gene Ontology Enrichment 解析	70
3.2.6	転写因子とモジュール間の制御関係の予測	70
3.2.7	ショウジョウバエ転写制御ネットワークとの比較解析及びネットワーク縮約	71
3.2.8	最上流転写因子と乾燥耐性関連遺伝子群間の制御関係の有無の妥当性検証	72
3.2.9	RNA-seq データの可用性	73
3.3	結果	73

3.3.1	トレハロース処理および再水和時系列遺伝子発現量の定量および発現変動遺伝子の検出	73
3.3.2	転写制御ネットワークの推定	77
3.3.3	協調発現モジュールの検出	82
3.3.4	転写制御ネットワークと協調発現モジュールの統合	86
3.3.5	ショウジョウバエ遺伝子制御ネットワークとの比較	89
3.3.6	NF-YC から乾燥耐性関連遺伝子群への制御関係に関する妥当性検証	93
3.4	考察	95
3.4.1	再水和直後におけるミトコンドリア転写活性と乾燥耐性機構の関係	95
3.4.2	GO 解析結果を利用したエフェクター遺伝子による新たな乾燥耐性機構発見の可能性	97
3.4.3	NF-YC による乾燥耐性制御の可能性	97
3.4.4	乾燥耐性を持たないヤモンユスリカが乾燥耐性を持たない理由	98
3.4.5	Positive Feedback Loop 構造を利用した乾燥耐性制御機構	98
第 4 章	結論	100
4.1	結言	100
4.2	Pv11 細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の同定	100
4.3	Pv11 細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定	101
4.4	今後の展望	102

目 次

1.1	本研究で対象とするアンハイドロビオシスの生命活動における位置 付け	4
1.2	乾燥耐性におけるアンハイドロビオシス誘導過程	5
1.3	水置換仮説概念図	7
1.4	乾燥の進行に伴うトレハロース産生量の違い	8
1.5	遺伝子制御の概念図	20
1.6	遺伝子制御ネットワークの構成	21
1.7	転写制御ネットワークに基づく肺がん悪性化転写因子 NFIX, PTCH の同定	22
1.8	Miloらによる種々の遺伝子制御ネットワークおよび電子回路におけ る Feed Forward Loop 構造に関する解析結果	23
1.9	遺伝子発現における Feed Forward Loop 制御による入出力関係	24
1.10	構造学習型転写制御ネットワーク推定の概念	27
1.11	回帰型転写制御ネットワーク推定の概念	31
2.1	各サンプル間における発現変動遺伝子に対する M-A Plot	44
2.2	各サンプルにおける遺伝子発現量 (TPM) に基づく階層的クラスタ リング結果	45
2.3	CAGE-seq および RT qPCR によって定量された Pv.07646 の発現量	48
2.4	GO:0008152 (metabolic process) および GO:0055114 (oxidation-reduction process) がアノテーションされた遺伝子に対する階層的クラスタリ ング結果	51
2.5	Pv.07646 と SNF2 histone linker PHD-ring helicase のアライメント 結果	54
2.6	Pv.07646 と DNA 修復遺伝子である RAD16 のアライメント結果	55
2.7	トレハロース処理時に顕著な発現上昇が確認されなかった乾燥耐性 関連遺伝子群の時系列発現	62
3.1	全サンプルにおける主成分分析の結果	75
3.2	転写因子及びそれ以外の遺伝子のトレハロース処理及び再水和にお ける発現変動遺伝子の数	76
3.3	推定したトレハロース処理時の転写制御ネットワーク	78

3.4	推定した再水和時の転写制御ネットワーク	79
3.5	推定したトレハロース処理及び再水和の統合転写制御ネットワーク	80
3.6	遺伝子制御における重要な構造の概念図と統合転写制御ネットワーク に存在する Feed Forward Loop (FFL) 及び Feedback Loop (FBL) 構造の個数	81
3.7	WGCNA により検出したモジュールの時系列発現量のヒートマップ	84
3.8	転写因子と協調発現モジュールの制御関係をまとめた遺伝子制御ネット ワーク	87
3.9	ショウジョウバエと Pv11 細胞の転写制御ネットワークの比較解析 結果	90
3.10	ショウジョウバエと Pv11 細胞の転写制御ネットワークの比較解析 結果から得られた縮約ネットワーク	91
3.11	最終縮約遺伝子制御ネットワーク	92
3.12	NF-YC を最上流とした乾燥耐性関連遺伝子群を直接制御する転写 因子 (PvG029178) への制御関係が少なくとも一つ存在する確率 . . .	94
3.13	各サンプルの染色体及びミトコンドリアゲノムへのマッピング率 . .	96
3.14	本研究で示唆した Pv11 細胞における乾燥耐性 ON/OFF 制御機構の 概念図	99
4.1	乾燥耐性システムの全容と現在明らかにした制御機構	105

表 目 次

1.1	乾燥耐性獲得時に働く乾燥耐性関連遺伝子とその生理活性に関する知見	12
1.2	様々な生物種において必要な乾燥耐性獲得期の長さ	14
2.1	CAGE-seq データに対する前処理及びマッピング結果	39
2.2	各サンプル間における発現変動遺伝子の個数	43
2.3	種々の DNA 修復機構に関わる遺伝子と配列相同性が確認された遺伝子に関する Pv11 細胞, ネムリユスリカおよびヤモンユスリカにおける発現上昇の有無	56
2.4	先行研究で示された乾燥耐性関連遺伝子の Pv11 細胞, ネムリユスリカおよびヤモンユスリカにおける発現変動遺伝子数	58
2.5	本研究で示唆した Pv11 細胞の乾燥耐性に関わる遺伝子に関する Pv11 細胞, ネムリユスリカおよびヤモンユスリカにおける発現上昇の有無	59
3.1	各サンプルにおけるマッピング結果	67
3.2	各乾燥耐性関連遺伝子ファミリーに対する, 最大所属モジュールと所属遺伝子	85
3.3	各モジュールを直接制御する転写因子及び HSF(PvG01882) による制御の有無	88

第1章 緒言

1.1 本論文の構成

生物はその生命活動に水を必須な物質として要求し、この水が失われる乾燥ストレスは多くの生物にとって致命的な要因の一つとなる。特に、陸生の生物は乾燥ストレスに常に晒されることから、これによる死を回避するためにその多くが進化の過程で乾燥ストレスを寛容化する分子機構を獲得してきた。特にこの分子機構の一つである無代謝状態 (Anhydrobiosis) への移行は、99%以上の生体内の水を失ったとしても水を再度与えることで元の生活環に戻る機構である [1]。これまでにワムシ (rotifers)、クマムシ (tardigrades)、線虫 (nematodes) や植物、そしてネムリユスリカ (*Polypedilum vanderplanki*) の幼虫など様々な生物がこの機構を持つことで乾燥ストレスによる死を寛容化することが知られている [2–6]。さらに2010年に樹立されたネムリユスリカの胚由来培養細胞である Pv11 細胞は事前に高濃度のトレハロースを含む培地で処理するトレハロース処理により、乾燥ストレスに晒されても水を与えることで細胞分裂を再開することが確認され、このことから乾燥耐性が細胞単位で達成される仕組みであることが示された [7]。一方で、この Pv11 細胞の乾燥耐性機構について分子レベルで解明した研究事例はなく、本機構は未知であった。本研究では、Pv11 細胞における乾燥耐性機構の解明を目指し、本機構を構成する遺伝子群およびそれらを制御する遺伝子制御ネットワークの推定を行なった。

第1章は緒言で、乾燥耐性を有する生物に関してこれまでに報告されてきた分子生物学的な知見について包括的に紹介する。そして、乾燥耐性機構解明に有用な遺伝子制御ネットワーク推定方法を概説し、次いで本研究の目的について述べる。

第2章では、トレハロース処理、乾燥、再水和時における Pv11 細胞の時系列 CAGE-seq データから、Pv11 細胞の乾燥耐性に関わる遺伝子群の推定に関して述べる。これらの遺伝子群の機能に基づき Pv11 細胞特有の乾燥耐性機構に関して得られた知見を報告する。

第3章では、第2章で得られた遺伝子群を含めてこれまでに報告されてきた乾燥耐性関連遺伝子群の発現を制御する遺伝子制御ネットワークの推定に関して述べる。推定した遺伝子制御ネットワークに基づき、Pv11 細胞の乾燥耐性関連遺伝子群の主要な制御を担う可能性のある転写因子やこれら遺伝子群の発現の ON/OFF を調節する可能性のある制御構造について報告する。

第4章では、明らかになった遺伝子制御ネットワークから考えられる Pv11 細胞

の乾燥耐性機構に関してまとめ、この推定した乾燥耐性機構の検証方法、そして乾燥耐性機構の制御による別生物種への乾燥耐性付与に関する今後の展望について議論する。

1.2 乾燥耐性

“(前略) 8月31日に、(とても暑い日が3日間も続いたので) 水は殆ど干上がってしまいましたので、樋の鉛の上にある泥を指で押さえても砂粒ほどの大きさの1滴にも満たないほどの水しかそこからは出てきませんでした。この水には透明な小動物がほんの僅か見付けられましたが、緑色や赤い色の小動物は皆死んでいました。(中略) 2月8日、あの材料を5ヶ月以上も戸棚の中のきれいな白い紙の上に置いておきました。その少量をきれいなガラス管に入れ、沸騰した雨水をさましてからそこに入れました。1時間半ほどして、すでに1匹の小動物が泳いでいるのを見ましたし、その他に、沢山のまだ丸くなったまものを見ました。そして3時間後には、これと同じ種類と違った小さな種類のを見ました。(後略)“

Antonie van Leeuwenhoek から
Hendrik van Bleyswijk への手紙(邦訳一部抜粋) [8,9]

1702年にAntonie van Leeuwenhoekが王立協会のHendrik van Bleyswijkに宛てた手紙で、樋に溜まった乾いた堆積物に存在した卵形の微視的な動物(後にワムシであることが明らかになる)に水を与えることで再び動き出すことを見つけたという言及が、乾燥しても死なず水を与えられると生き返る現象に関する最初の記述である[8,10]。その後、このような現象が乾燥以外のストレスに対しても起きる現象として多くの研究者によって発見され、この死を寛容する現象がクリプトビオシス(Cryptobiosis, 潜在的もしくは隠された生命活動)として定義された[11]。これらの生物種では代謝活性が測定不能になるほどに低下することが報告されている[1]。このクリプトビオシスは生物が受けるストレスとして分類されており、氷結ストレスによるクリオビオシス(Cryobiosis), 熱ストレスによるサーモビオシス(Thermobiosis), 酸素欠乏ストレスによるアノキシビオシス(Anoxybiosis), 浸透圧ストレスによるオスモビオシス(Osmobiosis), そして乾燥ストレスによるアンハイドロビオシス(Anhydrobiosis)がこれまでに報告されている(図1.1)[11]。これらの中でも乾燥ストレスによって誘導されるアンハイドロビオシスに関しては、99%の体内の水分¹を失っても再び水を与えると生命活動を再開するというような極限的な環境下で無代謝状態へ移行可能な生物種が多数確認されている[1]。この

¹体内の水分量は一般に含水率(Water Content, %H₂O)として定量される。含水率は $\%H_2O = \frac{W_t - W_d}{W_t} \times 100$ (W_t: 個体の質量, W_d: 個体の乾燥質量)で算出される[12]。この時、乾燥質量W_dは事前に個体から水分が検出されなくなる処理後における質量を計測しておく。例えば、クマムシの一種である*Macrobiotus areolatus*の場合には180°Cの加熱処理を3時間行うことで完全に水を失わせることができ、その時の質量W_dは12 μgと計測されている。

アンハイドロビオシスを誘導することができる機構を乾燥耐性と呼ぶ。これまでに極限的な環境下において乾燥耐性を持つ生物種としては、ワムシ、クマムシ、線虫、植物、そしてネムリユスリカが報告されてきた [2-6]。乾燥耐性を持つ生物種におけるアンハイドロビオシス誘導は主に、1) 体内の水分量が減少していく過程である乾燥耐性獲得期、2) 完全に代謝が停止する無代謝期、3) 水を与えられることで生体内の水分量が増加する再水和期に分けられる (図 1.2)。生と死という生物にとっての二つの状態とは一線を画したこの状態への移行および生への復帰を理解するために多くの分子生物学的研究が行なわれ、この現象をひき起こすエフェクター遺伝子と乾燥耐性に関わる生理学的な知見が蓄積してきた。さらに、2010 年には乾燥耐性を持つ動物由来の細胞として、ネムリユスリカの胚由来培養細胞である Pv11 細胞が世界で初めて樹立された。これにより、この非常に不可思議な現象は細胞単位で達成可能であることが示された [7]。

この乾燥耐性の機構を理解することは、前述の通り死とは異なる状態への移行による死の回避という生命にとっての根源的な問いを明らかにするという理学的な観点に留まらず、乾燥させられることにより個体が非常に軽量になることから物を輸送する際のコスト低減、保存が難しい生鮮食品などの長期保存を可能にする技術の開発など工学的な観点からも有意義と言える。特に Pv11 細胞のように、細胞単位で乾燥耐性が成り立つという知見は、この乾燥耐性をシンプルに別細胞へ付与できる可能性が高いことを示しており、Pv11 細胞の乾燥耐性機構の解明は前述の乾燥耐性を制御したいという工学的観点からも強く望まれていると言える。

本節では、これまでに乾燥耐性に関して報告されてきた分子生物学的な知見を乾燥耐性獲得期、無代謝期、再水和期における代謝、シグナル伝達、そして遺伝子制御の観点から説明し、現在得られている乾燥耐性機構の概要についてまとめる。そして、現在得られている乾燥耐性機構に関する知見のみでは乾燥耐性を制御するという目的を達成するためには不十分である点を指摘し、乾燥耐性機構を理解するためには、その制御機構の理解が必要不可欠であることを述べる。

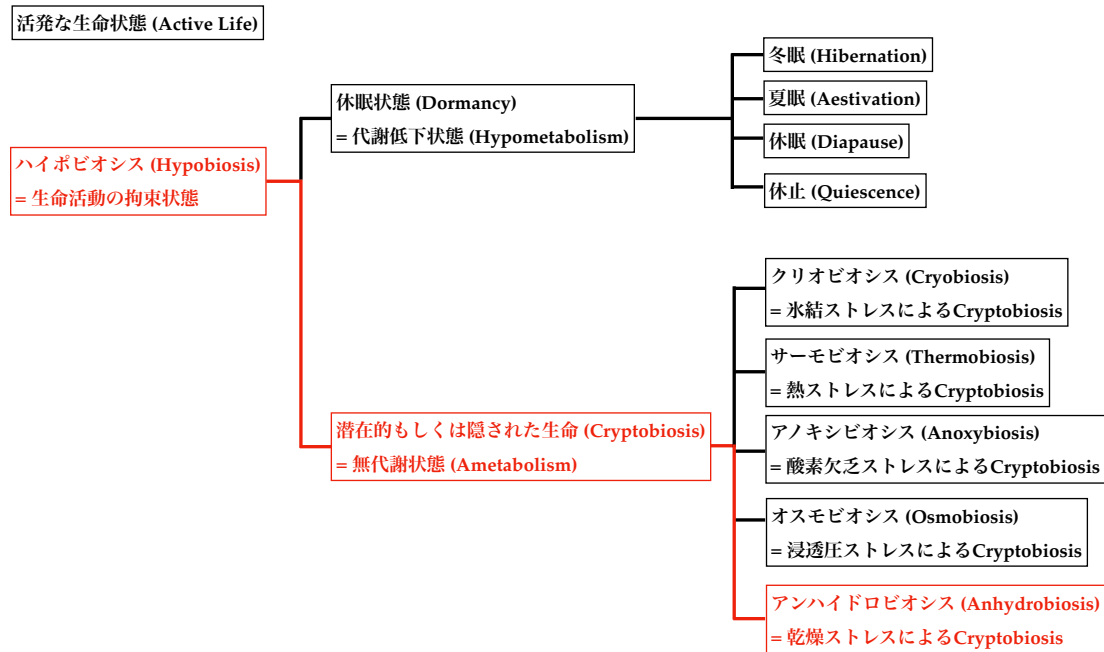


図 1.1: 本研究で対象とするアンハイドロビオシスの生命活動における位置付け

Keilin によって定義された生命活動の状態 [11] に基づくと、生命の状態は活発な生命状態とハイポビオシスと呼ばれる生命活動の拘束状態にまず分けられる。ハイポビオシスはさらに代謝活性の度合いで分類され、低代謝状態である休眠状態と無代謝状態であるクリプトビオシスに分けられる。クリプトビオシスはさらに受けるストレスによって、氷結ストレスによるクリオビオシス、熱ストレスによるサーモビオシス、酸素欠乏ストレスによるアノキシビオシス、浸透圧ストレスによるオスモビオシス、そして乾燥ストレスによるアンハイドロビオシスに分類される。本研究で対象とするアンハイドロビオシスは生命活動が拘束されている乾燥ストレスによって誘導される無代謝状態となる。

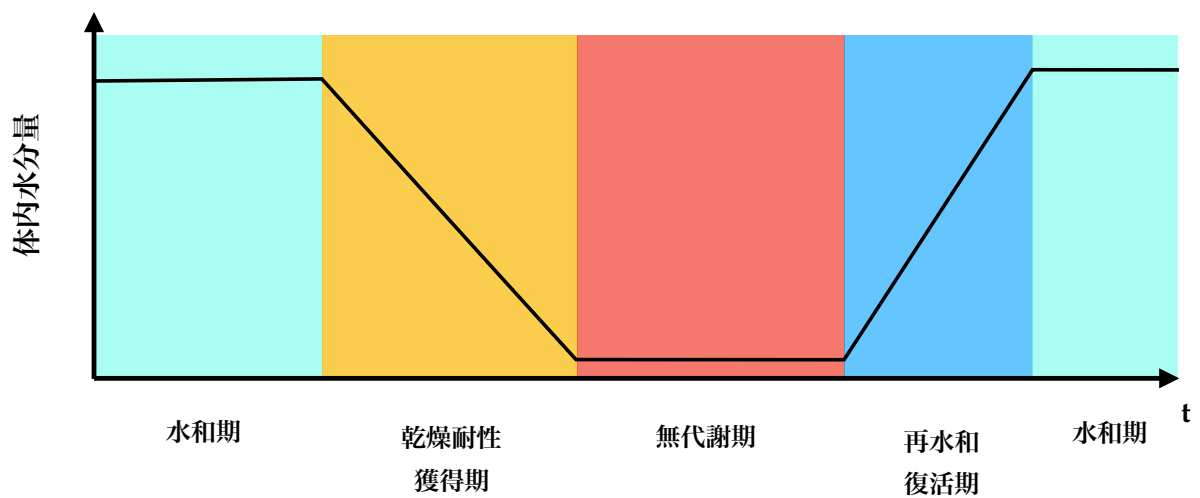


図 1.2: 乾燥耐性におけるアンハイドロビオシス誘導過程

一般的に乾燥耐性を持つ生物は体内の水分量が減少していく過程である乾燥耐性獲得期に、乾燥しても死なないために様々な分子を準備する。そして完全に代謝が停止する無代謝期においては顕著な分子種の変化が見られなくなる。そして水を与えられることで生体内の水分量は増加し、再水和期において代謝の再開など元の生活環を成り立たせる上で必要な分子の生成が行われる。

1.2.1 乾燥耐性獲得期

乾燥耐性獲得期とは、乾燥により体内の水分量が十分に下がりきるまでの間を指す。この期間は乾燥耐性を持つ生物にとって死を回避するための準備期間として考えられている。この期間において、乾燥耐性を持つ生物では通常とは異なる代謝、シグナル伝達、そして遺伝子発現が確認されている。実際に乾燥耐性に関する分子生物学的な知見の最も多くは、この乾燥耐性獲得期に関して蓄積している。本節ではこれらの反応経路ごとに乾燥耐性に関して報告されている分子生物学的な知見についてまとめ、最後に乾燥耐性獲得期に必要なとする時間の観点から、本研究で対象とする Pv11 細胞の乾燥耐性機構を明らかにする上で注目すべき反応経路について指摘する。

1.2.1.1 代謝

これまでに乾燥耐性を持つ生物種において広く、乾燥ストレス受容後の最初の応答として植物および動物においてはトレハロース、植物の種においてはスクロースといった糖の活発な代謝が確認されている [13, 14]。特に、トレハロースは無還元糖であることから細胞に対しての毒性はなく、細胞膜やタンパク質との間で乾燥時に欠乏する水を代替し水素結合を形成することで (水置換)、これらの物質の立体構造の安定化を促し乾燥による障害を寛容化している可能性が水置換仮説として報告されている (図 1.3) [1, 14-17]。また、トレハロースが乾燥耐性において必要な分子であることの証拠として、ネムリユスリカ幼虫では、緩やかな乾燥時には自重の 20% 以上になる質量のトレハロースを産生するのに対し、急激な乾燥時にはトレハロース産生が行えずに死亡するという報告がある (図 1.4) [6]。さらに、乾燥耐性を持つ Pv11 細胞はトレハロースの自発的な産生が行えないため、事前に高濃度のトレハロースを含む培地で処理するトレハロース処理を行わなければ、乾燥、再水和後の細胞分裂再開ができないことが報告されている [18]。このように、乾燥耐性を持つためには無還元糖であるトレハロースなどの産生が必要である。また、このトレハロースの産生を促すための酵素である trehalose-6-phosphate synthase (TPS) が乾燥に対して高発現することがネムリユスリカの幼虫で確認されており、乾燥に応じたトレハロース産生が活発になる理由についても報告されている [19]。しかしながら、TPS が乾燥に応じてなぜ発現上昇を示すか、といった TPS の発現と乾燥ストレスの間の関係はいまだ明らかにされておらず、代謝経路であるトレハロース産生を上流で制御するシグナル伝達や遺伝子制御についても明らかにされていない。また、乾燥耐性における代謝の関与として、ATP の AMP への変換が活発化するという報告がある [20]。この主な理由としては、乾燥時において要求されるエネルギーはほぼないため、この反応を進めることによって AMP としてエネルギー源を貯蔵しているのではないかと考えられている。

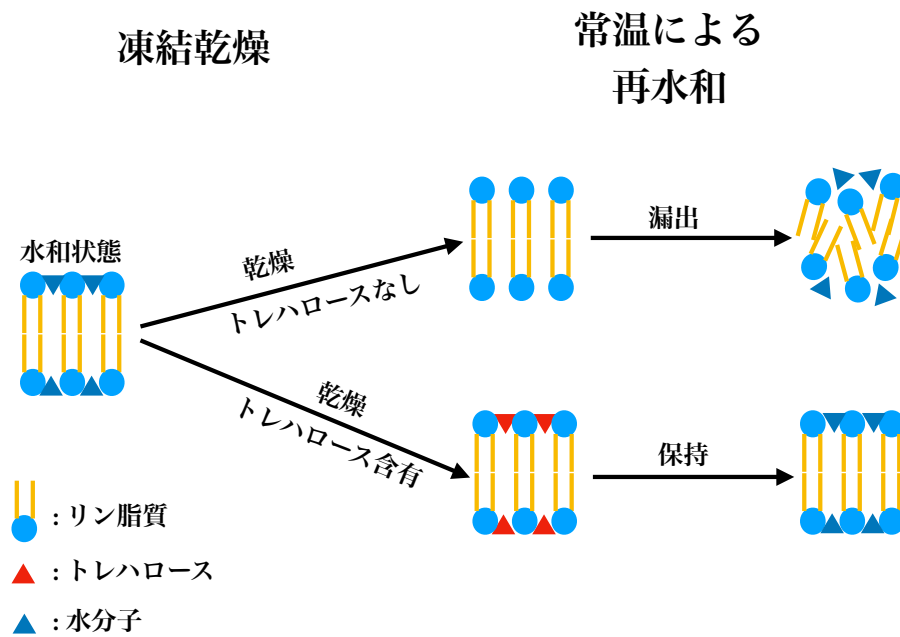


図 1.3: 水置換仮説概念図

乾燥時においてトレハロースが細胞膜近傍に存在することで、水を代替し細胞膜を保護する。これによって再水和後の細胞立体構造の変化による細胞内分子の漏出を防ぎ細胞を保護する。 [17] より改変, Adapted from [17]. ©1985, with permission from Elsevier.

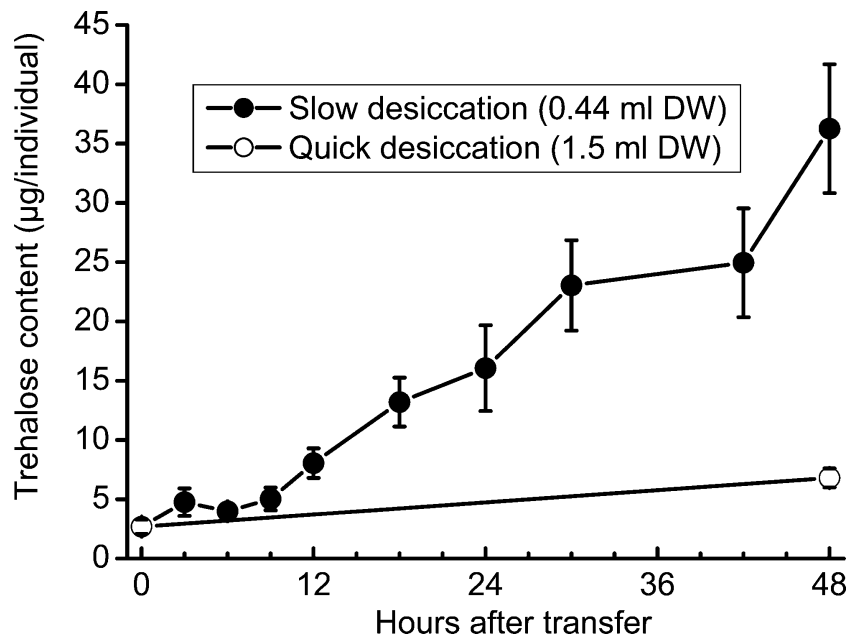


図 1.4: 乾燥の進行に伴うトレハロース産生量の違い

ネムリユスリカ幼虫の乾燥時トレハロース産生量を表す。横軸は乾燥させてからの時間を示し、縦軸は高速液体クロマトグラフで定量されたトレハロース量を示す。幼虫の乾燥方法はペトリ皿に 0.44 ml の水分を含ませたろ紙と 1.5 ml を含ませたろ紙に幼虫を置き、48 時間で乾燥するように蒸発させる (0.22 ml/day, 0.75 ml/day)。乾燥が急激に進むことを仮定している 1.5 ml 浸したろ紙に置いた場合においては 0.44 ml に浸したろ紙に置いた場合に対してトレハロースが産生されていないことが確認できる。各時刻におけるトレハロース産生量は 8 から 10 回の生物学的反復実験により定量され、その量は mean \pm SE で値を示す。Reprinted from [6]. ©2005, by permission of Oxford University Press.

1.2.1.2 シグナル伝達

これまでに乾燥耐性を持つ生物であるクマムシおよび線虫において、乾燥というストレスをどのようなシグナルとして受容されるのか、ということを開いた研究がある。乾燥に応じて応答するシグナル伝達経路としては、MAPK/ERK, JNK, p53 および AMPK 経路がある [21–23]。特に線虫 (*Aphelenchus avenae* および *Panagrolaimus superbus*) においては、MAPK/ERK, JNK, p53 に関してシグナル伝達経路を構成する分子のリン酸化が誘導されていることが確認されている [22]。これらのシグナル伝達経路が乾燥に応じて活性化する理由としては、乾燥に際する水欠乏により浸透圧が変化することを感受して活性化すると考えられている [24]。このように乾燥に応じたシグナル伝達経路の候補は多く確認されている一方で、これらシグナル伝達の末端に当たる遺伝子発現を誘導する転写因子や、ターゲットとなる遺伝子との関連性はいまだ報告がなく、前述の TPS の発現上昇との関係性についてもいまだ明らかではない。

1.2.1.3 遺伝子制御

これまでに乾燥耐性を持つ生物種において乾燥後に発現上昇を示す遺伝子群、およびその生理活性に関する知見は多い (表 1.1)。

まず、乾燥耐性を持つ生物種において、乾燥後に高発現が確認された遺伝子としては LEA (Late Embryogenesis Abundant) 遺伝子がある [3, 19, 25–27]。LEA 遺伝子が翻訳された LEA タンパク質は乾燥ストレスを受けていない状態ではランダムコイル形状を示し生理活性は見られない一方で [28]、乾燥ストレスを受けると立体構造は変化し、乾燥に応じた他タンパク質の変性を抑える分子シールドとしての役割を果たすことが報告されている [29]。このように、LEA 遺伝子は乾燥状態における生体内の物質の破壊を分子シールドとして防御することで乾燥耐性に関与している。

次に乾燥耐性において重要な役割を担うことが報告されている、乾燥時に発現上昇を示す遺伝子としてチオレドキシシン (Thioredoxin, TRX) がある [19, 30]。チオレドキシシン遺伝子が翻訳されたチオレドキシシンタンパク質は他タンパク質の酸化をシグナルとして、システイン残基のチオール・ジスルフィド反応を介して還元を促すことでタンパク質の酸化を防ぐ [31]。乾燥条件下においては、例えばネムリユスリカを例にとると、ガラス体 (トレハロースが不規則な空間配置を維持したまま固形化した状態) になると、活性酸素 (Reactive Oxygen Species, ROS) の生成における種々の分子の運動性は落ち、その結果化学反応量の増大、すなわち活性酸素量は増大する [32]。このように、チオレドキシシン遺伝子は乾燥状態の生体にとって障害となる活性酸素による影響を抑えることで乾燥耐性に関与している。

次に乾燥耐性への関与が確認されている遺伝子としては protein-L-isoaspartate (D-aspartate) O-methyltransferase (PIMT) が報告されている [19, 33]。PIMT 遺伝子が翻訳されたタンパク質は、他タンパク質における L および D 体のアスパラギ

ン酸の障害を修復することが報告されている [34]. 乾燥時における特異的なアスパラギン酸の障害に関する報告はないものの, PIMT の発現上昇が乾燥時特異的であることから, この遺伝子は乾燥時におけるタンパク質への障害を修復することで乾燥耐性への関与があると考えられている.

次に乾燥時に高発現が確認された乾燥耐性関連遺伝子としてヘモグロビン (Hemoglobin, HB) がある [19]. ヘモグロビンの一般的な生理機能としては, 血中内の O_2 含有量を増加させ全身に輸送する働きがある [35]. 乾燥耐性への関与としては, この生理機能を利用して乾燥条件下の個体中において酸素枯渇に対する対策として酸素の貯蔵を行うことで乾燥耐性に寄与しているという知見がある [19].

次に乾燥時に高発現が確認されている乾燥耐性関連遺伝子にアクアポリン (AQP) がある [19,36]. アクアポリンは細胞膜のリン脂質二重膜における浸透圧を制御している [37]. 前節でも述べたように, 乾燥耐性における乾燥ストレスの受容には浸透圧の変化が重要な役割を担っている可能性が示されている. このことと併せて考えると, アクアポリンの乾燥時における高発現は, 浸透圧変化を感受するためと考えられる.

最後に乾燥耐性への関与が報告されている遺伝子としては, 乾燥時のトレハロース産生に関わる trehalose-6-phosphate synthase (TPS) がある [19]. TPS は UDP-glucose と Glucose-6-phosphate から Trehalose-6-phosphate を生成する反応を触媒する酵素である [38]. Trehalose-6-phosphate からトレハロースを生成する反応を触媒する酵素としては trehalose-6-phosphate phosphatase (TPP) があるが, 乾燥時における高発現はいまだ確認されておらず, 当該反応は乾燥時特異的な発現上昇を伴わない可能性がある [19].

ここまでで述べてきたように, 乾燥耐性を成り立たせる上で必要な遺伝子と, その翻訳後のタンパク質の生理活性に関しては非常に広範に研究が進められてきた. 一方で, これらの遺伝子の高発現を誘導する遺伝子制御機構については長らく未知の状態だった. これに対して 2018 年に Pavel Mazin らによって, これら乾燥耐性に関与する遺伝子群のエンハンサーサイトの網羅的な解析が行われ, 熱ショック因子 (Heat Shock Factor, HSF) の結合サイトがこれらの領域に豊富に存在することが示された [39]. さらに彼らは, 熱ショック因子を Pv11 細胞において knock down し, 野生型の Pv11 細胞との間で上述の乾燥耐性関連遺伝子の発現を比較したところ, LEA 遺伝子およびチオレドキシシン遺伝子の一部, そして TPP の発現が有意に減少したことを報告した. これにより熱ショック因子が乾燥耐性において必要な遺伝子制御を担っていることが示されたと言える. これは一見, 乾燥耐性を支配する転写因子とその遺伝子制御の全貌が明らかになったように見えるが, 実際には熱ショック因子は乾燥耐性を持たないショウジョウバエなどの生物にもその存在は知られてもいる [40]. さらに, Mazin らの報告では, 上述した遺伝子以外の乾燥耐性関連遺伝子の発現は変動しなかったことや, 熱ショック因子の knock down では再水和後の蘇生率を 0% にできなかったことが示されている. これらの事実から熱ショック因子は乾燥耐性において必要な転写因子ではあるが, これだけ

では乾燥耐性機構を支えるのには十分とは言えず，乾燥耐性を持つ生物特有の他の転写因子による転写制御が存在することを示していると言える。

表 1.1: 乾燥耐性獲得時に働く乾燥耐性関連遺伝子とその生理活性に関する知見

乾燥耐性関連遺伝子名	略記名	生理活性
Late Embryogenesis Abundant	LEA	乾燥状態で他タンパク質の変性を抑える
チオレドキシシン	TRX	乾燥状態における生体にとっての障害となる活性酸素による障害を抑える
protein-L-isoaspartate (D-aspartate) O-methyltransferases	PIMT	乾燥時におけるタンパク質への障害を修復する
ヘモグロビン	HB	乾燥条件下において酸素枯渇に対する対策として酸素の貯蔵を行う
アクアポリン	AQP	乾燥によるシグナルを受け取るための浸透圧変化を感受する
trehalose-6-phosphate synthase	TPS	トレハロースの前段階代謝物質である Trehalose-6-phosphate の合成を触媒する
熱ショック因子	HSF	LEA 遺伝子およびチオレドキシシン遺伝子の一部, そして TPP の発現を制御する

1.2.1.4 乾燥耐性獲得期のまとめ

本節では、乾燥耐性獲得期に関与する代謝、シグナル伝達、そして遺伝子制御に関して先行研究を述べた。このように様々な反応経路を利用することで乾燥による死を回避することができる乾燥耐性を持つ生物種ではあっても、生体内の水分を急速に失う乾燥には耐えきれず、個体としての死を回避できないことが一般的である [6, 18, 41–44]。さらに乾燥耐性を誘導するために必要な乾燥耐性獲得期の長さは生物種によって大きな乖離がある (表 1.2)。一般に、代謝、シグナル伝達、遺伝子制御にかかる時間のオーダーは、それぞれ秒、十秒、分から時間オーダーであることが報告されていることを考慮すると [45–47]、乾燥耐性を持つ生物種ごとにその最も重要な反応経路は異なる可能性を示している。本研究で対象とする Pv11 細胞は乾燥耐性獲得期に 48 時間という時間オーダーの反応を要求することを考えると、Pv11 細胞の乾燥耐性獲得期におけるメカニズムを明らかにする上では遺伝子制御に注目することが有用であると考えられる。

表 1.2: 様々な生物種において必要²な乾燥耐性獲得期の長さ

生物名 (和名)	生物種名	必要な乾燥耐性獲得期の長さ	参考文献
クマムシ	<i>Ramazzottius varieornatus</i>	30 分	[41]
線虫	<i>Rotylenchulus reniformis</i>	1 時間	[42]
ワムシ	<i>Macrotrachela quadricornifera</i>	12 時間	[43]
ワムシ	<i>Philodina roseola</i>	12 時間	[43]
ワムシ	<i>Adineta oculata</i>	12 時間	[43]
クマムシ	<i>Hypsibius dujardini</i>	24 時間	[44]
ネムリユスリカ	<i>Polypedilum vanderplanki</i>	48 時間	[6]
Pv11 細胞	<i>Polypedilum vanderplanki</i>	48 時間	[18]

²再水和後の蘇生率がほぼ 100%を維持することができることを指す。

1.2.2 無代謝期

無代謝期とは乾燥耐性を持つ個体において水分量が十分に減少しきった状態における過程を指す。この期間は乾燥耐性を持つ生物における生命活動の停止期間と考えられている。無代謝期においては、その名称通りに個体における前述したトレハロースやスクロース以外の代謝活性は観測できない [1]。また、この過程において主要な役割を担うシグナル伝達や遺伝子発現制御に関する知見もなく、この過程においてこれらの何らかの制御機構が必要か否かについては明らかになっていない。一方で、これまでに無代謝期の可能持続時間が多くの生物において調査され、短いものでワムシ (*Macrotrachela musculosa*) では2~3ヶ月という報告から線虫 (*Tylenchus polyhynchus*) では39年間と、その間には非常に大きな乖離が存在する [10]。この個体毎の無代謝期の可能持続時間の違いは、乾燥耐性獲得期における水分量が減少していく過程にかける時間の長さに依存するらしい [48]。すなわち、乾燥される過程で水分量を十分に保持できず、乾燥時に生じる様々な障害を十分に寛容化できない状態で無代謝期に入らざるを得ないような生物種においては、無代謝期の可能持続時間を十分に長く取ることができないのではないかと考えられている。

1.2.3 再水和期

再水和期とは無代謝期の状態から水を与えられ個体としての活動を再開、または細胞分裂が再開するまでの過程を指す。この期間は乾燥耐性を持つ生物における生命活動の再開を行う期間として考えられている。再水和期においては、通常的生活環を成り立たせるための代謝の活性、活発なシグナル伝達および遺伝子発現がこれまでに広く確認されている。本節ではこれらの反応経路ごとに再水和期に関して報告されている分子生物学的な知見についてまとめ、最後に再水和期にかかる時間の観点から、本研究で対象とする Pv11 細胞の乾燥耐性機構を明らかにする上で注目すべき反応経路について指摘する。

1.2.3.1 代謝

再水和期における代謝に関しては、まずトレハロースやスクロースなど乾燥耐性獲得期に準備した代謝物が分解されることが報告されている [49,50]。また、乾燥時に蓄積した AMP を ATP に変換する反応が誘導される [20]。そして、最終的には炭水化物代謝 (Carbohydrate metabolism) が再開し、通常状態と変わらない代謝状態に戻ることが確認されている [51]。このように再水和期における代謝は基本的に、乾燥耐性獲得期に準備した物質の分解や、停止させた代謝経路の復活など元の生活環に戻るために必要な反応で占められていることが実験的に確認されている。

1.2.3.2 シグナル伝達

再水和期におけるシグナル伝達としては、乾燥耐性獲得期と同様に MAPK/ERK, JNK, p53 経路の活性化が確認されている [52]. 末端の遺伝子発現までは確認されていないが、これらのシグナル伝達経路の再水和に伴う活性化は通常の MAPK 経路が持つ細胞分裂など生命現象を再開させる上で必要な遺伝子発現の準備に重要であると考えられている. このように再水和期におけるシグナル伝達は代謝と同様に、通常の状態に戻る上で必要な遺伝子発現を駆動するための反応経路が活性化することが実験的に確認されている.

1.2.3.3 遺伝子制御

再水和期における遺伝子発現制御は、代謝、シグナル伝達において見られた通常状態への復帰に加えて、乾燥時に生じた障害を修復するための遺伝子発現が報告されてきた. その主な遺伝子としては抗酸化を促すチオレドキシシンと DNA の修復を行う遺伝子群が挙げられる [53]. 前述した通り、乾燥時には活性酸素が多量に生成され、タンパク質の酸化や DNA への障害が頻発する [54]. これらの乾燥時に蓄積した障害に対して、再水和時においては上述した遺伝子の高発現を誘導することによって対処する遺伝子制御系が存在する可能性がある.

1.2.3.4 再水和期のまとめ

ここまでで示した通り再水和期の主な現象としては、元の生活環に戻る上で必須な代謝活性の回復、それを成り立たせるためのシグナル伝達や遺伝子制御がある. さらに遺伝子制御においては特異的に乾燥時において蓄積した障害を修復する遺伝子発現を誘導することが確認されている. 乾燥耐性獲得期と同様に、再水和期に要する時間の観点から本研究で対象とする Pv11 細胞において重要となる反応経路に関して言及する. これまでにクマムシや線虫、ネムリユスリカの幼虫においてはおよそ 1 時間以内に活発な行動の再開が確認されている [19, 42, 55]. 一方で Pv11 細胞の細胞分裂の再開は 3 日程度経たないと確認されない [18]. このことは細胞レベルでの乾燥耐性においては、細胞分裂の再開までに非常に長い時間を要することを意味しており、前述した一反応に分から時間オーダーの時間を要求する遺伝子制御が律速として働いている可能性を示している. このことから、Pv11 細胞の再水和期におけるメカニズムを明らかにする上では遺伝子制御に着目することが有用であると考えられる.

1.2.4 乾燥耐性に関するまとめ

本節では乾燥耐性に関してこれまでに得られた分子生物学的な知見について、乾燥耐性獲得期、無代謝期および再水和期に分類して包括的に述べた. このように、

乾燥耐性について分子生物学的な知見は広く報告されている一方で、これらのほぼ全てはクマムシや線虫、ネムリユスリカの幼虫といった個体での報告に限られており、Pv11細胞において報告された事実は代謝に関してのトレハロースの自発的な産生が不可能な点と、熱ショック因子が乾燥耐性における必要な遺伝子制御を担っている点のみになる。すなわち、細胞単位の乾燥耐性に関してはいまだ分子生物学的な知見はほぼないと言っても過言ではない。さらに、前述した乾燥耐性関連遺伝子群の制御機構に関しては個体、細胞問わず、いまだ知見がほぼない。

この細胞単位の乾燥耐性を明らかにする上で最初に注目すべき反応経路として、乾燥耐性獲得期および再水和期にPv11細胞が要する時間の観点から、本研究では遺伝子制御に着目した。次節では遺伝子制御の分子生物学的な位置付けを説明し、多数の遺伝子制御によって構成される遺伝子制御ネットワークを定義付ける。そして、遺伝子制御ネットワークを推定するための方法論についてまとめる。

1.3 遺伝子制御ネットワーク

1.3.1 遺伝子制御ネットワークの定義と研究事例

Pv11 細胞の乾燥耐性を明らかにする上で、前節でも述べたように律速段階となる遺伝子制御にまず注目する。遺伝子制御とは、ある転写因子が対象となるゲノム上の結合領域への結合を介して結合領域下流にある遺伝子の転写を制御する反応を指す (図 1.5)。この反応を多数の転写因子間およびターゲットとなる遺伝子間の反応として捉え、転写因子や遺伝子を頂点 (ノード)、間の反応を辺 (エッジ) として表したものを遺伝子制御ネットワークと呼ぶ (図 1.6) [56]。遺伝子制御ネットワークは転写因子と転写因子の制御関係のみで構成される転写制御ネットワーク、および転写因子とターゲットとなる転写因子以外の遺伝子との間の制御関係である末端遺伝子制御ネットワークから構成される [57] (しばしば、これら全てを包括した遺伝子制御ネットワーク自体を転写制御ネットワークと呼ぶ場合もあるが、本論文では前述の呼称で統一する)。遺伝子制御は遺伝子発現を記述し、シグナル伝達経路の構成要素であるタンパク質や代謝反応を触媒する酵素の発現制御に関わるため、細胞内において起きる現象を実質的に全て担っていると見なされている [45]。これまでに様々な生命現象を担う遺伝子制御ネットワークに関する研究が進んでおり、ヒト、大腸菌、酵母、線虫、ショウジョウバエやシロイヌナズナといったモデル生物に関しては、その遺伝子制御ネットワークが調べられ、対象とする生命現象との関連性が議論されている [58–62]。

遺伝子制御ネットワークの同定後にまず行うことができる対象とする生命現象に対するアプローチとしては、対象とする生命現象に関わる全遺伝子発現を制御する最上流の転写因子であるマスター転写因子の同定が挙げられる。先行研究では肺癌悪性化における転写制御ネットワークの同定および最上流の転写因子に注目することで、NFIX および PTCH1 が肺癌悪性化を担うマスター転写因子であることを同定し、肺癌治療における新たな創薬ターゲットの発見に成功したことを報告している (図 1.7) [63]。このように最上流の転写因子に着目することは生命現象のキーとなる転写因子の発見には有用であるが、これまでにトランスクリプトーム解析で一般的に行われてきた発現変動遺伝子解析でもこのような転写因子の検出は可能であり、この解析だけでは生命現象を明らかにする上で本質的な遺伝子制御ネットワークを明らかにする意義はないと言える。

そこでこれまでに遺伝子制御ネットワークと生命現象を結びつける試みが多数行われてきた。遺伝子制御ネットワークと生命現象を結びつける方法論としては、まず遺伝子制御ネットワーク全体としての特徴であるスモールワールド性やスケールフリー性を持つかどうかの検証やハブと呼ばれる多くの制御関係を持つ遺伝子に着目する大局的な特徴に注目する方法がある [64, 65]。Nabil Guelzim らは、出芽酵母の転写制御ネットワークがランダムに生成されたネットワークよりも平均経路長 (任意の 2 ノード間のエッジ数) が小さいというスモールワールド性に加え

てネットワークの直径 (任意の2ノード間のエッジ数のうち最大の値) が非常に小さいことを明らかにし, 出芽酵母はこのような転写制御ネットワークを持つことで遺伝子制御における情報伝達の遅れを減らしていることを示唆した [64]. また Reka Albert は, 様々な生物における転写制御ネットワークが出次数 (あるノードが別のノードに対して伸ばすエッジの数) 分布において, 一部のノードのみ大多数の出次数を持つというスケールフリー性を持つことを示し, このことから転写制御はこうした大多数の出次数を持つハブと呼ばれる転写因子によって全体の発現パターンが制御されることを明らかにした [65]. これらのように, 大局的な特徴に注目する方法論は遺伝子制御ネットワーク全体としての傾向を理解する上では重要な知見を提供するが, 具体的にどのような制御関係に基づき対象とする生命現象が制御されているかは明らかにすることができない. これに対して, ランダムには起き得ない頻出する構造として定義されるネットワークモチーフと呼ばれる局所的な特徴に注目する方法論が存在する [66]. Milo らはこのネットワークモチーフに着目することで, 大腸菌および酵母の遺伝子制御ネットワークにおいてフィードフォワードループ (Feed Forward Loop) と呼ばれる3遺伝子から構成される構造が, 3遺伝子間で構成可能な全遺伝子制御構造のうち唯一ネットワークモチーフとして検出されることを明らかにし, さらにこれまでに作成された電子回路においてもこのフィードフォワードループがネットワークモチーフとして検出されることを明らかにした (図1.8). 彼らはフィードフォワードループ構造が一過的な入力をカットして出力に伝えず, 十分な入力が入った時のみ出力に伝えるノイズ除去機構を備えているという知見と合わせて (図1.9) [67], 生物の転写制御が電子回路の構造と類似な, 高度なノイズ除去を行える制御によって成り立つ可能性を示した. このようにネットワークモチーフに基づく局所的な特徴に注目することで生命現象の制御に関する知見を得ることが可能になる. 本研究でも乾燥耐性の制御に向けた現象の理解を最終目標に掲げているため, この局所的特徴に注目した解析に基づくアプローチを取る.

しかしながら, この考え方を適用するためには遺伝子制御ネットワークが明らかになっている必要がある. これまでに Pv11 細胞を初めとして, 乾燥耐性に関する遺伝子制御ネットワークを明らかにした事例はない. そのため Pv11 細胞の乾燥耐性制御機構を明らかにする上で, まず乾燥耐性を形成する遺伝子制御ネットワークを推定する必要がある. 次節から, 遺伝子制御ネットワークを構成する転写制御ネットワークおよび末端遺伝子制御ネットワークの推定方法でこれまでに提案された方法論について概説し, それらの中で本研究の目的を達成する上で有用な方法論について対象とする転写因子数や実験的なコストの観点から議論する.

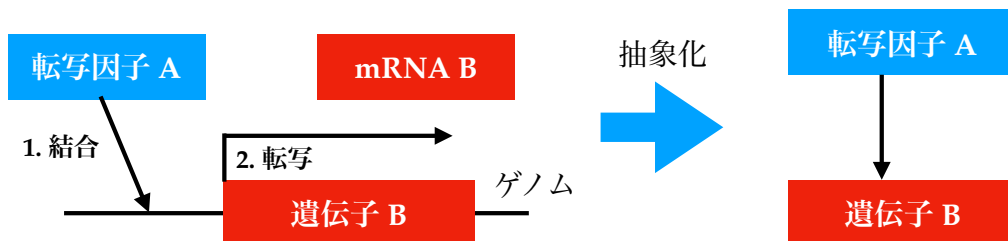


図 1.5: 遺伝子制御の概念図

遺伝子制御とは 1) 転写因子が対象となるゲノム上の DNA 結合領域へ結合する, 2) 結合領域近傍の遺伝子の転写を活性化することを総称したものであり, 遺伝子制御関係を表す場合には, 制御する転写因子と制御される遺伝子をそれぞれ頂点 (ノード) で表し, 制御関係を辺 (エッジ) で表す.

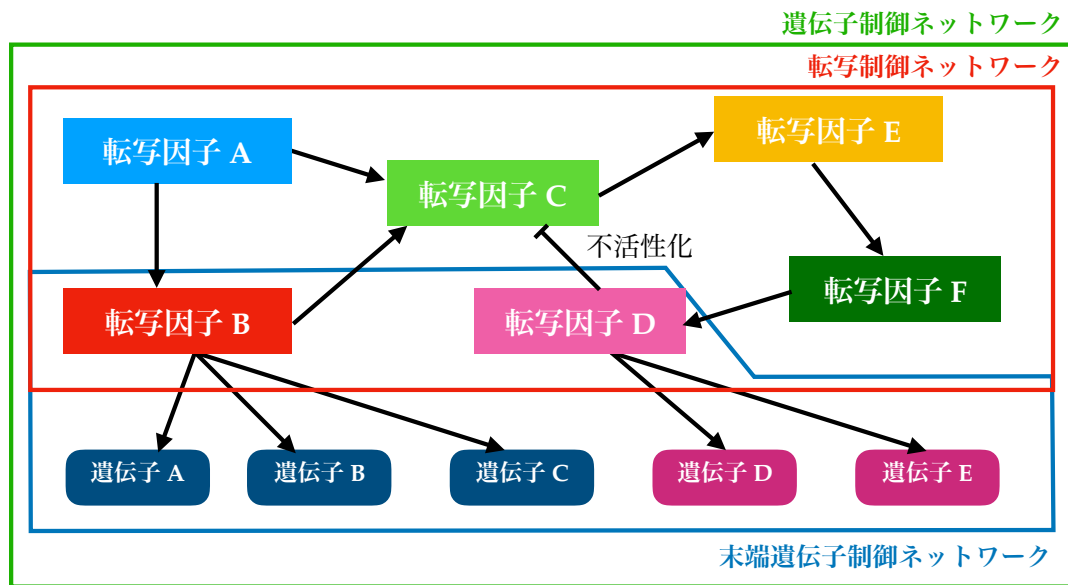


図 1.6: 遺伝子制御ネットワークの構成

遺伝子制御ネットワークは転写因子と転写因子の制御関係のみで構成される転写制御ネットワークと転写因子とターゲットとなる転写因子以外の遺伝子との間の制御関係である末端遺伝子制御ネットワークから構成される。

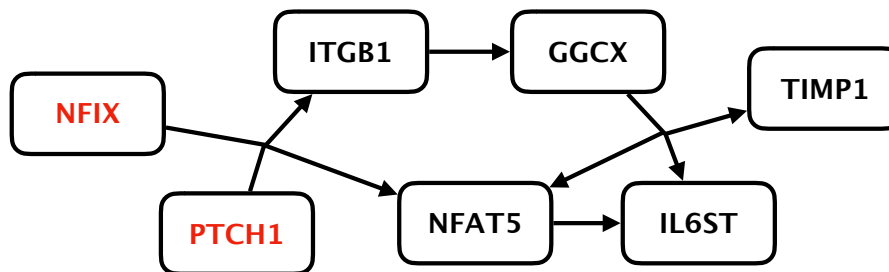
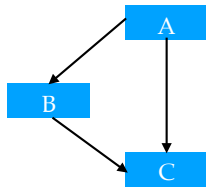


図 1.7: 転写制御ネットワークに基づく肺癌悪性化転写因子 NFIX, PTCH の同定

Rahman らは肺癌悪性化に関わる転写制御ネットワークを同定し, その最上流の転写因子に着目することで NFIX および PTCH がマスター転写因子として働くことを示唆した. [63] より改変. Adapted from [63], ©2017 Rahman, Abdul Murad, Mollah, Jamal and Harun, <http://creativecommons.org/licenses/by/4.0/>



Feed Forward Loop (FFL)

対象	対象種	N_{real}	$N_{rand} \pm SD$	Z-score
遺伝子制御 ネットワーク	大腸菌	40	7 ± 3	10
	出芽酵母	70	11 ± 4	14
電子回路 (Forward Logic Chip)	s1580	424	2 ± 2	285
	s38584	413	10 ± 3	120
	s38417	612	3 ± 2	400
	s9234	211	2 ± 1	140
	s13207	403	2 ± 1	225

N_{real} : FFLの個数

N_{rand} : ランダムに生成したネットワークにおけるFFLの個数

図 1.8: Milo らによる種々の遺伝子制御ネットワークおよび電子回路における Feed Forward Loop 構造に関する解析結果

Milo らは大腸菌および出芽酵母で既に同定されていた遺伝子制御ネットワークおよび電子回路 (フォワード方式の論理回路) において Feed Forward Loop (FFL) 構造の個数 (N_{real}) を算出した [66]. また, それぞれのネットワークにおけるノード数とエッジ数を保存したランダムなネットワークを 1, 000 個生成し, FFL 構造の個数 (N_{rand}) とその標準偏差 (SD) を算出した. また, 実際のネットワークの FFL 構造の個数の統計的有意性を検証するために $Z\text{-score} = (N_{real} - N_{rand}) / SD$ を算出した. その結果, 遺伝子制御ネットワーク, 電子回路ともに FFL 構造の個数はランダムなネットワークにおける個数よりも有意に多いことが明らかになった.

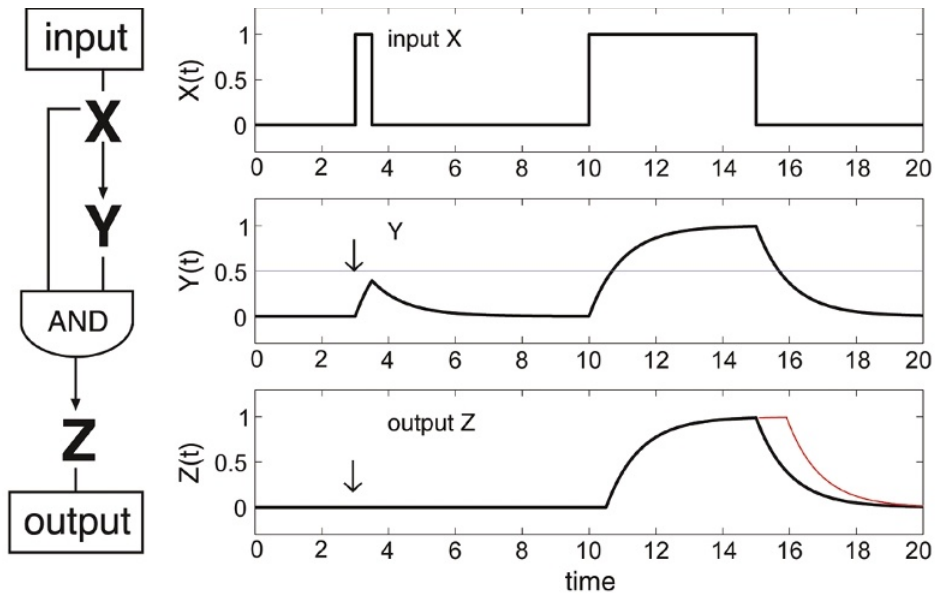


図 1.9: 遺伝子発現における Feed Forward Loop 制御による入出力関係

図は Feed Forward Loop 制御 (FFL) による遺伝子発現制御をモデル化した場合の時間に対する X , Y , Z の量を示す。これにより入力パルス状の場合には下流に入力が入らず、十分な時間入力が入った場合のみ下流を活性化するという性質があることが示された。このことから Feed Forward Loop 構造を持つことは遺伝子発現制御において一過的な入力をパスして持続的な入力だけを通す、ノイズ除去を行う制御機構であることが示唆された。また、 $X \rightarrow Y \rightarrow Z$ のみの反応の場合には Z の発現変動が遅れることが確認されており (赤線)、Feed Forward Loop 制御は持続的な入力に対してより速い応答をすることが示された。Reprinted by permission from Springer Nature Customer Service Centre GmbH: Nature Genetics [67], ©2002. Print rights of the final author's accepted manuscript for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/)

1.3.2 転写制御ネットワーク推定

転写制御関係の推定は一般的に、転写因子の発現変動を原因、その下流の転写因子の発現変動を結果とした因果を推定する方法に集約される。そのため同時刻における遺伝子発現のデータからでは、(相関関係しか導けないため) 転写制御関係を推定することは困難である。そこで、転写制御関係の推定には各転写因子の発現変動のタイミングの前後関係を利用して推定する時系列モデルが提案されている。時系列モデルによる推定手法は問題の定式化方法により大きく2つの手法に分類される。1つ目の手法は、与えられたデータに基づき転写制御構造にスコアを与えてスコアを最大化する転写制御構造を推定する構造学習型転写制御ネットワーク推定法である。2つ目の手法は、対象とする遺伝子発現を直前の別の遺伝子発現から回帰することで回帰モデルを得て、そのモデルの特徴量解析に基づき転写制御構造を推定する回帰型転写制御ネットワーク推定法である。本節では、構造学習型転写制御ネットワーク推定と回帰型転写制御ネットワーク推定の原理について概説し、本研究で用いるべき手法を対象とする転写因子数の観点から議論する。

1.3.2.1 構造学習型転写制御ネットワーク推定

構造学習型転写制御ネットワーク推定は、与えられたデータに基づき転写制御構造に対してスコアを算出し、そのスコアを最大化する転写制御構造を求めることでこれを同定する手法である (図 1.10)。数学的に定式化すると以下の式で表される。

$$G = \arg \max_{G'} (\text{Score}(G'|D)) \quad (1.1)$$

上式の G は求めたい転写制御構造、 D は与えられたデータ、 $\text{Score}(G'|D)$ は与えられたデータ D に基づく転写制御構造 G' のスコアを表す。一般に本問題を解く際には $\arg \max$ 関数の対象となる G' の候補は、転写因子の個数に対して組合せ爆発を起こすため、最適化手法で代替する近似計算が行われる。上記の事項に基づき、構造学習型転写制御ネットワーク推定手法では 1) Score の与え方、2) 最適化手法の選び方により様々な手法が提案されている。Score の与え方に関しては、データに対する構造の事後確率を求めるベイズ情報量規準 (Bayesian Information Criterion) や BD スコア、BDe スコア [68]、遺伝子発現の前後関係に基づくシャノンエントロピー [69] などが提案されている。このスコアの与え方により一般に、ベイズ統計に基づく手法をベイジアンネットワーク型転写制御ネットワーク推定、情報量に基づく手法を情報理論型転写制御ネットワーク推定と呼ばれる。また、最適化手法に関しては転写制御構造自体が互いに不連続な関係にあるため、一般には非勾配法と呼ばれる最適化手法に分類される貪欲法や焼きなまし法、遺伝的アルゴリズムなどが利用される [68]。構造学習型転写制御ネットワーク推定は原理的には $\arg \max$ 関数を正しく導くことができれば、データに適合する、すなわち現象

を再現する最適な転写制御ネットワークを推定することが可能である。しかし上述したように転写因子の増加に対して探索すべき転写制御構造 G' のパターンは爆発的に増加する。具体的には、対象とする転写因子の個数が n 個の場合、探索すべき G' の個数である $num_{G'}(n)$ は以下の数式に基づき増大する。

$$num_{G'}(n) = 4^{nC_2} \quad (1.2)$$

上式は、ある任意の二つの転写因子 A, B の間に、A から B への制御関係、B から A への制御関係、両方向の制御関係、制御関係なし、の 4 パターンが存在することから導くことができる。上式より明らかのように、探索すべき G' の個数は転写因子の数に対して指数関数的に増加し、例えば転写因子数が 10 個の場合には約 1.23×10^{27} 個、100 個の場合には約 10^{2475} 個となり、最適化手法を用いたとしても現状の計算機的能力では十分な探索を行うことができない。このように構造学習型転写制御ネットワーク推定手法は原理的に多数の転写因子を対象とした推定には向いておらず、近年の次世代シーケンサの発達に基づく網羅的な遺伝子発現量の定量化技術との相性が悪い。そこで、このような組合せ爆発の問題を解消する手法として提案された回帰型転写制御ネットワーク推定手法について次節にて紹介する。

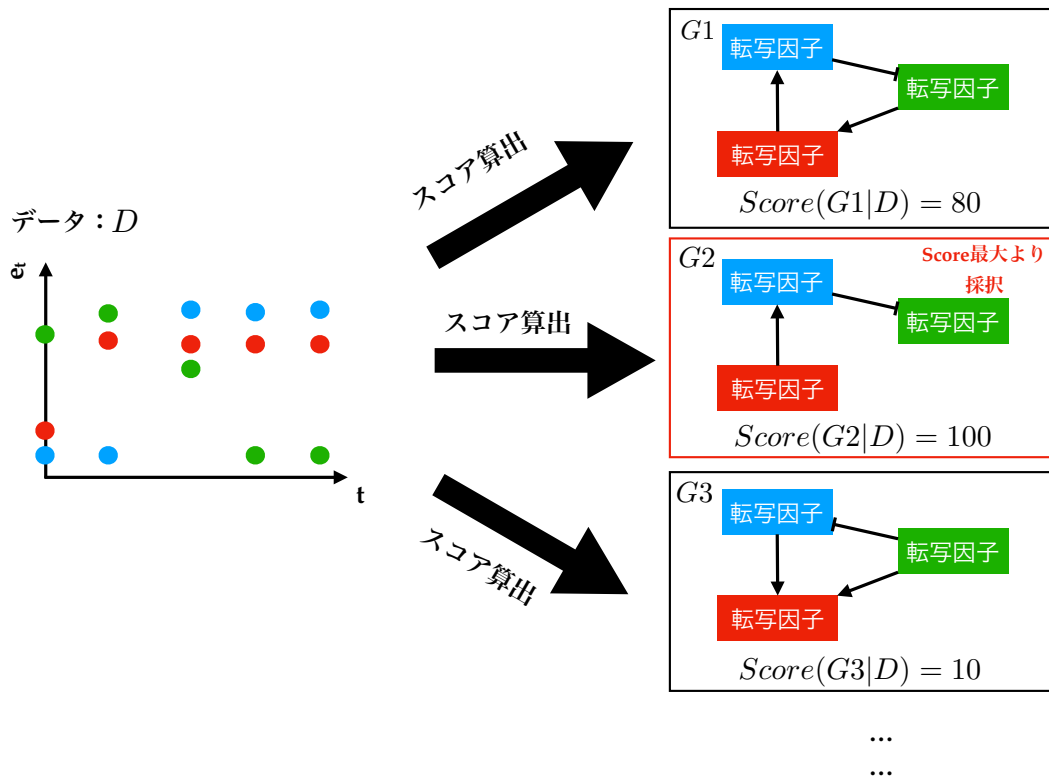


図 1.10: 構造学習型転写制御ネットワーク推定概念

構造学習型転写制御ネットワーク推定では、各転写因子の時系列発現量データ D が与えられた時の転写制御ネットワーク構造に対するスコアを算出し、そのスコアが最大になる構造を採択することで転写制御ネットワークを推定する。しかし、一般には探索すべきネットワーク構造の個数が転写因子数に対して指数関数的に増加するため、ヒューリスティックな最適化法を用いてスコアの良い構造を探索する。

1.3.2.2 回帰型転写制御ネットワーク推定

回帰型転写制御ネットワーク推定は構造学習型転写制御ネットワーク推定で問題となった組合せ爆発を回避する目的で提案されたネットワーク推定手法である [70]. 回帰型転写制御ネットワーク推定手法では、まず以下のモデルの回帰を行う [71].

$$e_t^i = f_i(e_{t-1}^{-i}) + \epsilon_{t-1}, \forall t \quad (1.3)$$

ここで、 e_t^i は任意の転写因子 i の時刻 t での発現量、 e_{t-1}^{-i} は転写因子 i 以外、すなわち転写因子 i を制御する可能性のある候補転写因子の時刻 $t-1$ での発現量、 ϵ_{t-1} は時刻 $t-1$ での誤差項、 f_i は e_t^i の発現量を説明するための関数である。本モデルのデータへの適合、すなわち回帰を行なった後に f_i の関数に基づき、 e_{t-1}^{-i} のうちの転写因子が関数 f_i への影響が大きかったかを解析することで、転写因子 i への制御関係を推定する (図 1.11)。これを全転写因子に対して行い、最終的には各転写因子からの制御関係の影響力で制御関係にランク付けを行い、閾値処理をすることによって転写制御ネットワークを推定する手法になる。本手法は f の関数形の設定の仕方によって、 f の関数形を e_{t-1}^i に対して線形な式や遺伝子発現の揺らぎをより詳細にモデル化した式に固定するモデル固定型の手法 [72, 73] と f の関数形を固定せずに機械学習に基づき回帰を行うモデルフリー型手法 [71, 74, 75] に分類することができる。回帰型転写制御ネットワーク推定は上述の通り、構造学習型転写制御ネットワーク推定において問題となった組合せ爆発の問題を、 n 回の回帰の問題で解くこと (n regression problem) で解消することができている。この組合せ爆発問題の解消がなされた回帰型転写制御ネットワーク推定は、ハイスループットな遺伝子発現定量技術との相性が非常に良い。そのため近年では f の関数形の設定方法や回帰モデルからの特徴解析に関して研究の進捗が著しく、多数のアルゴリズムが提案されている。

その中でも、現在様々な遺伝子制御ネットワーク推定のベンチマークの問題に対して最大の精度を示す推定アルゴリズムとして BTNET が提案された [71]. BTNET はモデルフリー型の推定アルゴリズムであり、機械学習アルゴリズムとして勾配ブースティング木を採用している [76]. 勾配ブースティング木は木型回帰アルゴリズムの一種であり、誤差関数の勾配を元に最適化を行う機械学習アルゴリズムである。木型回帰アルゴリズムは、被説明変数 Y に対して説明変数 $X(x \in X)$ のうちで閾値処理をして Y を分割した際に分割された Y の値のそれぞれの分散が最も小さくなるような x とその閾値を選択する。これを分割された Y に対しても同様に行い続け、最終的に分割された Y の平均値を出力値として得る方法である (この時、ここまでの操作で作成された説明変数で分割し続けるものを“木”と呼ぶ)。

勾配ブースティング木では、上記の方法で出力値の決定を行なった後に、実際に説明したい Y の値との誤差を算出し、誤差が減るように出力値の補正を行い、繰り返し木を生成することで、よりデータに適合した回帰を達成する手法である。以下に勾配ブースティング木を利用した BTNET のアルゴリズムにおける手順をま

とめる。但し、以下においては f を推定したい関数、 Y_i をある時刻 i における制御対象の転写因子の発現量、 \mathbf{x}_i をある時刻 i における対象とする転写因子以外の発現量を示すベクトル、 N を全時系列数、 L を誤差関数、 β を学習係数、 T を学習回数とする。

- (a) 各時刻の推定したい関数の出力値を全時系列データの平均を用いて $f_0 = \bar{Y}$ として設定する
- (b) N 個のデータを用いて f の関数値に適合するように関数 f を勾配ブースティング木で学習させる
- (c) 適合したいデータ Y と関数の出力値 $f(x)$ を用いて誤差関数 $L(f(x_i), Y_i)$ ($= \frac{1}{2}(Y - f(x))^2$) を元に各時刻における関数出力値の一階微分 $R(f(x_i), Y_i) = \frac{\partial L(f(x_i), Y_i)}{\partial f(x_i)} = Y_i - f(x_i)$ を算出する
- (d) 各時刻の関数出力値を $f_{t,i} = f_{t-1,i} + \beta \times R(f(x_i), Y_i)$ で算出する
- (e) 上記 (b) から (d) を学習回数 T に達するまで繰り返し行う

ここまでのステップで制御対象の転写因子の発現量を別の転写因子の一時刻前の発現量で説明することができる関数 f を得ることができる。続いて、このデータに適合した f に基づき制御関係の推定を行う。この際に最も重要なこととしては、勾配ブースティング木において選択されたある転写因子 x によって分割された Y の発現量の分散が分割される前に比べてより小さくすることができるのであれば、そのような x は Y への制御関係の確からしさは高いという点になる。この概念は基本的には、ある変数を説明変数として加えた際に被説明変数に関するモデルの説明性が、その変数を説明変数に加えない場合に比べて顕著に上がること (より正確に述べると、最小二乗 (Ordinary Least Squares, OLS) 残差が有意に減少すること) がその説明変数と被説明変数の間に因果があるとする Granger 因果と同義である [77]。このような因果の概念に基づき、BTNET では対象となる転写因子 Y に対するある転写因子 g による制御関係の確からしさ $VIS_ensemble(g)$ を f を構成する勾配ブースティング木を用いて以下の通りに定義した。

$$VIS_ensemble(g) = \frac{1}{T} \sum_{t=1}^T VIS_t(g) \quad (1.4)$$

$$\begin{aligned} \text{ただし, } VIS_t(g) = & \sum_{S_i \in S, \text{ split variable of } S_i \text{ is } g} |S_i| \times Var(S_i) \\ & - |S_{i, \text{ left}}| \times Var(S_{i, \text{ left}}) - |S_{i, \text{ right}}| \times Var(S_{i, \text{ right}}) \end{aligned} \quad (1.5)$$

上式の S は Y の要素によって構成される集合を表し、 $|S|$ は S の要素数を表す。また $Var(S)$ は S の構成要素の分散を示し、 $S_{i, \text{ left}}$ および $S_{i, \text{ right}}$ は S の集合を説明変数である転写因子 g の発現量で閾値処理をして分割した Y の部分集合を表す。

(1.5) 式より明らかに，分割前の S の分散に対して分割後の分散が小さくなればなるほど， $VIS_t(g)$ は大きくなり，結果的に $VIS_{ensemble}(g)$ も大きくなる．すなわち，この指標の大小により転写因子 g による被制御転写因子 Y への制御関係の確からしさ (Granger 因果の度合い) を考えることができる．転写制御ネットワークを構成する全ての g そして被制御転写因子 Y についても全ての g について算出することにより，2つの転写因子間の制御関係の $VIS_{ensemble}$ 列を得ることができ，最終的にこの $VIS_{ensemble}$ の値に対して適切な閾値を設定することで転写制御ネットワークを推定することが可能になる．先行研究では上記のアルゴリズムに基づき様々な転写制御ネットワーク推定に関するベンチマークの問題に対して，閾値処理による真の制御関係と偽の制御関係を純粋に分割できたかどうかを測る指標である AUPR (Area Under Precision and Recall) と AUROC (Area under Receiver Operator Curve) が他アルゴリズムに対して最も高いことが示された [71].

このように，回帰型転写制御ネットワーク推定手法は構造学習型転写制御ネットワーク推定では不可能な多数の転写因子に対する制御関係の推定を可能にする．本研究で対象とする Pv11 細胞のリファレンスゲノムであるネムリユスリカのゲノムには 176 個の転写因子が確認されている [39]. この転写因子の個数に対して転写制御ネットワークを推定するためには回帰型転写制御ネットワーク推定手法が有用であり，先行研究において最大の精度が示されている BTNET を本研究では採用した．

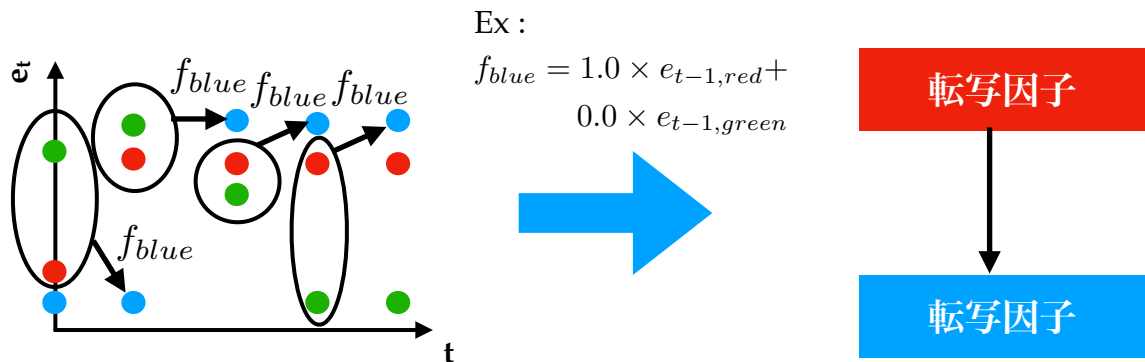


図 1.11: 回帰型転写制御ネットワーク推定のご概念

回帰型転写制御ネットワーク推定では、まず制御対象となる転写因子を定め(青色転写因子)、その転写因子の発現変化を別の転写因子の一時刻前の発現量で回帰する。その後、データに適合した回帰モデルを用いて、制御対象となる転写因子を説明できる転写因子を解析する。図では、回帰モデルを線形モデルで仮定した場合の例を示す。回帰モデルが $f_{blue} = 1.0 \times e_{t-1, red} + 0.0 \times e_{t-1, green}$ として得られたとすると、この場合は赤色転写因子の発現量によって青色転写因子の一時刻後の発現量が良く説明できることになるため、赤色転写因子から青色転写因子への制御関係が予測される。これを制御対象となる転写因子を全転写因子に対してそれぞれ独立に行うことによって、転写因子の個数が n 個の場合、 n 回の回帰と n 回の回帰モデルの解析で転写制御ネットワークを推定することができる。

1.3.3 末端遺伝子制御ネットワーク推定

末端遺伝子制御ネットワークとは転写因子と転写因子以外の遺伝子の間に存在する遺伝子制御関係を指す。転写因子以外の遺伝子を要素に含むため、全遺伝子が別の遺伝子を制御し得ることを仮定する転写制御ネットワーク推定の方法論は末端遺伝子制御ネットワークの推定には用いることができない。そのため、末端遺伝子制御ネットワーク推定には別の概念に基づく方法論が提案されている。それは、転写因子の下流遺伝子群のプロモータ領域への結合を推定することで遺伝子制御関係を推定することである。本節ではこの概念に基づく推定方法である ChIP-seq データを用いた手法と転写因子の結合配列モチーフデータを用いた手法について説明する。

1.3.3.1 ChIP-seq データを用いた末端遺伝子制御ネットワーク推定

ChIP-seq (Chromatin Immunoprecipitation followed by sequencing) では免疫沈降法を利用して対象とするタンパク質が結合した DNA 配列を選択的に読み出す [78]。ChIP-seq ではまず、対象とする細胞をホルムアルデヒドで処理することで、タンパク質と DNA を強固に架橋する。その後、細胞に対して超音波処理を施すことで細胞を破碎、および DNA を 200 から 600 塩基の範囲で分割する。そして、結合領域を明らかにしたいタンパク質の抗体を用いることで対象となるタンパク質が結合した DNA 配列のみを免疫沈降法で取得する。この取得したタンパク質と DNA の複合物を超音波で切断し、得られた DNA 配列をシーケンサーで読み出し、リファレンスゲノムにマッピングすることで注目したタンパク質の結合配列を網羅的に得ることができる。このタンパク質を転写因子として ChIP-seq を行い、結合配列近傍の遺伝子を同定することにより転写因子が制御し得るターゲット遺伝子を同定することが可能になる [79]。ChIP-seq は対象とする転写因子と DNA 結合配列の関係を網羅的に探索できることから、転写制御ネットワークの同定に用いられた事例がある [61]。一方で、ChIP-seq は、1) 転写因子に対する抗体がなければ行うことができない、2) 上流である転写因子に関する網羅性はない(一つずつ抗体を作成して上記の手順を踏む必要があるため)、という 2 点から十分に実験環境が整っている生物種以外には適用しづらいという問題点がある。さらに ChIP-seq によって明らかにできることは転写因子が対象とする配列に結合するか否かという点のみであるため、その結合が下流遺伝子の発現を誘導するかどうかは明らかにできない。そのため、遺伝子制御を明らかにする上では結合配列を同定できる ChIP-seq と下流遺伝子の発現の有無を明らかにすることができる RNA-seq を併用して行うことが一般的である [61]。本研究で対象とする Pv11 細胞における転写因子については、いまだ抗体作成ができた転写因子は存在しない。そのため、この ChIP-seq を用いた末端遺伝子制御ネットワークの推定は現状不可能である。このような実験的なアプローチによる末端遺伝子制御ネットワーク推定に代替する方

法論として、次節では転写因子の結合配列モチーフデータを用いた手法について述べる。

1.3.3.2 結合配列モチーフデータを用いた末端遺伝子制御ネットワーク推定

転写因子は一般に 10 nt 程度の配列を認識して結合し遺伝子発現を制御する [80]. この配列は転写因子毎に異なることが報告されており、このような配列のことを転写因子に対する結合配列モチーフと呼ぶ [81]. これまでに、様々な生物種において転写因子と結合配列モチーフの関係が ChIP-seq の実験結果などを元にまとめられており、データベースも拡充している [82–84]. また、これらデータベースにおいては結合配列は一意的な塩基配列では定義されておらず、配列位置ごとの A, T, G, C の塩基確率を表す Position Weight Matrix (PWM) として定義されている. 本節で説明する手法は、この結合配列モチーフデータを利用することで実験的に得られる ChIP-seq に基づく結合配列を代替し転写因子と末端遺伝子群間の制御関係を推定する. 結合配列モチーフデータを用いた末端遺伝子制御ネットワークの推定においては、基本的に 1 つの転写因子対 1 つの遺伝子間での制御関係ではなく、1 つの転写因子対多数の遺伝子間における制御関係を推定することが一般的である [85]. このような手法をモチーフエンリッチメント解析 (Motif Enrichment Analysis, MEA) と呼ぶ. 本節ではモチーフエンリッチメント解析で最高の精度を示す CLOVER アルゴリズムについて説明する [85, 86]. ある転写因子の結合配列モチーフと結合可否を考える配列 L が与えられた時に、結合に関する尤度 $LR1(L)$ を以下のように定義する.

$$LR1(L) = \prod_{k=1}^W \frac{q(k, L_k)}{p(L_k)} \quad (1.6)$$

ここで、 W は結合配列モチーフの塩基長、 L_k は L 内の k 番目の塩基、 $q(k, L_k)$ は対象としている転写因子の結合配列モチーフにおける k 番目の L_k の塩基の確率、 $p(L_k)$ はランダムな配列における L_k の塩基の確率を表す. ランダムな配列から求められる塩基の確率は一般的に、結合を判定する生物種のゲノムを構成する塩基の割合で代替することが行われる. この $LR1(L)$ は値が高いほど、対象とする転写因子の PWM に基づく結合配列モチーフと類似した配列であることを示す. 続いて、この $LR1(L)$ を用いて結合を判定する遺伝子上流領域 S への結合に関する尤度 $LR2(S)$ を以下のように定義する.

$$LR2(S) = \frac{1}{M_S} \sum_{L \in S} LR1(L) \quad (1.7)$$

ここで、 M_S は上流領域 S における結合モチーフ配列検証箇所の個数を示し、 $M_S = |S| - W + 1$ である ($|S|$ は S の配列長を示す). 上式は上流領域 S 全域における $LR1(L)$ の平均を表しており、値が高いほど上流領域には転写因子の結合配列モ

モチーフと類似した配列が存在することを表す. 続いて対象とする遺伝子群の上流領域全体を通じた結合配列モチーフの存在に関する尤度を定義する. この際に, 全ての遺伝子群の上流領域に必ず結合領域が存在するという仮定は厳しすぎるため, 遺伝子群の数 N に対して i 個の上流領域における結合配列モチーフの存在に関する尤度として以下のように $LR3(i)$ を定義する.

$$LR3(i) = \frac{1}{{}_N C_i} \sum_A \prod_{S \in A} LR2(S) \quad (1.8)$$

ここで, A は対象とする遺伝子群の上流領域 S について i 個分取り出して作ることができる全ての部分集合を要素としてもつ集合であり, その要素数は ${}_N C_i$ である. 上記の $LR3(i)$ は一部の領域 S でも潤沢に結合配列モチーフを持っていれば値は高いため, 前述した全ての遺伝子群の上流領域に必ず結合領域が存在しなければいけないという仮定を緩和している. 最後にこの $LR3(i)$ を用いて, i を 1 から対象とする上流領域の個数である N までの平均を取ることで, ある遺伝子集団の上流領域に対する転写因子の結合配列モチーフの存在に関する尤度 $LR4$ を以下のように定義する.

$$LR4 = \frac{1}{N} \sum_{i=1}^N LR3(i) \quad (1.9)$$

CLOVER ではこの $LR4$ の自然対数をとった $\ln(LR4)$ を転写因子の対象とする遺伝子群上流領域への結合可能性のスコアとして用いている. 続いて, こうして得られた $\ln(LR4)$ が統計的に有意に高いかどうかを検討するために, CLOVER では遺伝子上流領域 S の配列をランダムに入れ替えて $\ln(LR4)$ を算出する処理を複数回³行うことで, 本来は結合配列がないはずの上流領域群に対する $\ln(LR4)$ の分布を取得する. これを帰無仮説に従う際の統計量 $\ln(LR4)$ の分布として考えることで, 実際に得られた $\ln(LR4)$ の値よりも大きな値を算出した処理の回数を総処理数で除算した値を p 値として得る. 最終的にこの p 値を有意水準 α で閾値処理することにより, 対象とする転写因子の遺伝子集団上流領域への結合の有無を判定する. このような手法によって ChIP-seq のデータなしに転写因子の結合配列の情報があれば末端遺伝子制御ネットワークの推定ができる. Pv11 細胞の転写因子に関しては結合配列モチーフの情報はいまだ明らかにされていないが, ネムリユスリカの近縁種であるショウジョウバエの転写因子に関しては既に多くの転写因子の結合配列モチーフが同定されている [82–84]. 本研究ではこれを代用することにより ChIP-seq のデータなしにモチーフエンリッチメント解析を用いた末端遺伝子制御ネットワークの推定を試みた. この際にはモチーフエンリッチメント解析において最高の精度が示されている CLOVER を解析手法として採用した.

³回数を増やすほど真の帰無仮説に従う分布が生成でき, CLOVER では 1000 回で十分としてこの値をデフォルト値として採用している

1.3.4 遺伝子制御ネットワーク推定に関するまとめ

本節では遺伝子制御ネットワークを構成する転写制御ネットワークおよび末端遺伝子制御ネットワークの推定方法に関してこれまでに報告されてきた手法についてまとめた。転写制御ネットワーク推定は構造学習型転写制御ネットワーク推定法と回帰型転写制御ネットワーク推定法が提案されており、本研究で対象とする Pv11 細胞の転写因子の個数の観点から回帰型転写制御ネットワーク推定手法が有用であることを指摘した。末端遺伝子制御ネットワーク推定には ChIP-seq を利用した実験的に転写因子の結合に関するデータを利用した推定手法と結合配列モチーフと呼ばれる転写因子の結合配列に関する情報を元に転写因子とターゲット遺伝子群の制御関係を推定する手法についてまとめ、本研究で対象とする Pv11 細胞の転写因子はまだ ChIP-seq で必要な抗体の作成が行えていない点から、後者の手法を用いて末端遺伝子制御ネットワークを推定することが有用であることを述べた。

1.4 本研究の目的

1.2 節では、乾燥しても死なず水を与えられると生き返る機構である乾燥耐性についての定義付けを行い、これまでの研究で明らかにされた乾燥耐性機構に関する代謝、シグナル伝達および遺伝子制御に関して乾燥耐性獲得期、無代謝期および再水和期について説明した。これまでにクマムシや線虫、ネムリユスリカの幼虫といった個体レベルの乾燥耐性機構に関しては多くの分子生物学的な知見が蓄積している一方で、Pv11 細胞における細胞レベルの乾燥耐性機構に関してはいまだ明らかになっていない点が多い。さらに個体、細胞を問わず乾燥耐性に関わる遺伝子群を制御する機構に関してはほぼ未知の状態にある。本研究では Pv11 細胞の乾燥耐性制御機構の解明を目指して、Pv11 細胞の乾燥耐性に関わる種々の現象に要求する時間の観点から遺伝子制御にまず着目し、Pv11 細胞の乾燥耐性に関わる遺伝子群の同定およびそれら遺伝子群の制御機構を明らかにすることを目的とする。

本研究ではまず、Pv11 細胞の乾燥耐性獲得期、無代謝期および再水和期における CAGE-seq データを用いることで、各過程における発現変動変動遺伝子を検出する。これら遺伝子群の機能解析および 1.2.1.3 小々節および 1.2.3.3 小々節でまとめた個体における乾燥耐性関連遺伝子と比較することにより Pv11 細胞特異的な乾燥耐性機構について考察する。

続いて Pv11 細胞の乾燥耐性関連遺伝子の制御機構の解明を目指して 1.3.2.2 小々節および 1.3.3.2 小々節でまとめた方法論を用いた転写制御ネットワーク推定および末端遺伝子制御ネットワーク推定に基づく遺伝子制御ネットワーク推定を行う。得られた遺伝子制御ネットワークに基づき、1.3.1 小節で述べたように最上流転写因子の同定とネットワークの局所的な特徴であるネットワークモチーフに注目す

ることで、Pv11細胞の乾燥耐性がどのような制御機構に基づき達成されているかを考察する。

本研究ではこれらの結果を総括して、今後の乾燥耐性機構の制御による別生物種への乾燥耐性付与を視野に入れた乾燥耐性制御機構に関して得られた知見をまとめる。

第2章 Pv11細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の推定

2.1 背景

生命にとって必要不可欠な水を失う乾燥ストレスは多くの生物種にとって致死的なストレスである。このストレスに対して、ワムシやクマムシ、線虫、植物及びネムリユスリカの幼虫はアンハイドロビオシスと呼ばれる無代謝状態に入ることによって死を回避しており、99%以上の水を失っても生き残ることができる(1.2節参照)。このような乾燥耐性の獲得に対しては、水置換仮説と呼ばれるトレハロースのような水の代替物となれる化合物が細胞膜や生体内物質を保護することで、乾燥による生体の死の回避に貢献していることが提唱されている(1.2.1.1小々節参照)。また生体内の水を失う乾燥耐性獲得期においては、乾燥時に発生する生体にとっての障害となる影響を除去するLEA遺伝子やチオレドキシソニン遺伝子、PIMT遺伝子や乾燥時に枯渇する酸素の保持のためのヘモグロビン遺伝子、浸透圧変化の受容に寄与する可能性のあるアクアポリン遺伝子の高発現が確認されている(1.2.1.3小節参照)。こうしたトレハロースの生成や乾燥耐性関連遺伝子の高発現を行なった個体は、水を与えられることで元の生活環に戻ることができる。

2010年に、この機構を細胞単位で持つことができるネムリユスリカ胚由来培養細胞であるPv11細胞が樹立された[7]。Pv11細胞は600 mMのトレハロースを含む培地に48時間浸す、トレハロース処理を施すことで乾燥耐性を獲得する。この状態から12日間の乾燥後に(RH < 10%)再び培地を与える再水和処理後1時間で、16%の細胞が死を回避した。その後3日程度経過後に細胞分裂の再開が確認された[18]。また同様の実験手順を施した他の昆虫細胞系列(Sf9, BmN-4, AeAl-2, AnCu-35およびS2)では乾燥を経た後の再水和後1時間における生細胞が確認されなかったことから、Pv11細胞は昆虫細胞において唯一乾燥耐性を持ち、再水和による細胞分裂再開が可能な昆虫細胞であることが示された。しかしながら、Pv11細胞に対する網羅的なトランスクリプトーム解析はいまだ行われておらず、この細胞がトレハロース処理、乾燥及び再水和を経て元の状態に戻っているのか明らかにされていない。さらに、Pv11細胞の乾燥耐性の獲得及び再水和による細胞分裂再開に寄与する遺伝子及びその生化学的なメカニズムはいまだ明らかとなっていない。

いない。

これらの未知な点を明らかにするために、本研究では Pv11 細胞に対するトレハロース処理、乾燥及び再水和時に発現する mRNA を網羅的に解析するために Pv11 細胞に対する当該処理時における CAGE-seq データを解析した。この時系列 CAGE-seq データに対して発現変動遺伝子解析を行い、得られた発現変動遺伝子にどのような生化学的な機能を持つ遺伝子が含まれているかを調べるために GO (Gene Ontology) 解析を行なった。本研究では Pv11 細胞に対するトレハロース処理、乾燥及び再水和時の遺伝子発現を網羅的に解析することにより、トレハロース処理、乾燥時には生体にとって障害となる物質を除去する遺伝子が、再水和時には乾燥時に壊れた DNA を修復する遺伝子が主要な役割を担う可能性を示した。

2.2 方法

2.2.1 CAGE-seq データの前処理

本研究では Pv11 細胞に対するトレハロース処理、乾燥および再水和時における CAGE-seq データを農業・食品産業技術総合研究機構の黄川田研究室より提供いただいた。まず提供いただいたデータの概要について説明する。Pv11 細胞は 10% FBS を含む IPL-41 培地で培養され (T0), その後 600 mM のトレハロースと 10% の FBS を含む IPL-41 培地で 48 時間培養するトレハロース処理をした (T48)。その後この試料は 8 日および 10 日間、乾燥器において RH < 10% の条件で乾燥された (D8, D10d)。10 日間の乾燥処理後、試料に 10% FBS を含む IPL-41 培地で 3 時間および 24 時間再水和した (R3, R24)。これら T0, T48, D8, D10d, R3, R24 のサンプルは全て生物学的反復数 3 で準備された。これらのサンプルに対して RNAiso-Plus kit (TaKaRa, 滋賀, 日本) を用いて total RNA が採取され、NucleSpin RNA kit (Macherey-Nagel, Düren, ドイツ) で浄化された。RNA のクオリティーは Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) で検証され、Qubit 2.0 fluorometer (Thermo Fisher, Waltham, MA, USA) で定量された。その後、5' 末端に Cap がついた RNA のみ CAGE バーコードタグをつけて 1 本鎖 cDNA に合成された。得られた CAGE タグがついた cDNA は HiSeq-2500 (Illumina, San Diego, CA, USA) でシーケンスされた。ここまでの処理は黄川田研究室にて実施された実験概要となる。その後、得られた各リードは 1) Cap 構造の可能性のある最初の 3 塩基の除去, 2) 3' 末端から 7 塩基 A が続いている場合に PolyA の可能性を考慮して当該塩基を除去, 3) 2 つ以上の N 塩基を含むリードを除去した。この処理後に得られたリードは Bioconductor の QuasR パッケージ [87] の qAlign 関数を用いてネムリユスリカゲノム (midgBase: <http://150.26.71.110/midgbase/>) にマッピングした (表 2.1)。マッピングされたリードは同パッケージの qCount 関数を用いてカウントし、サンプル毎の各遺伝子のマッピングされたリード数をカウントした。

表 2.1: CAGE-seq データに対する前処理及びマッピング結果

totalSequences: シーケンスされた全リード数, tooShort: 14塩基以下のリード数, tooManyN: 2つ以上のN塩基を持つリード数, totalPassed: totalSequences から tooShort 及び tooManyN のリード数を引いたリード数, mapped: マッピングされたリード数, unmapped: マッピングされなかったリード数, mapping rate: totalPassed リード数に対する mapped リード数の割合

	前処理結果				マッピング結果			
	totalSequences	tooShort	tooManyN	totalPassed	mapped	unmapped	MappingRate	
T0-Sample1	9329152	1089685	33	8239434	1682490	6556944	20.4	
T0-Sample2	8618767	1642930	23	6975814	2463740	4512074	35.3	
T0-Sample3	5486398	1002516	12	4483870	1517136	2966734	33.8	
T48-Sample1	8249645	907419	27	7342199	1282918	6059281	17.5	
T48-Sample2	4527055	923602	9	3603444	1340121	2263323	37.2	
T48-Sample3	5862177	1172255	20	4689902	1747773	2942129	37.3	
D8-Sample1	7162127	867901	23	6294203	1263235	5030968	20.1	
D8-Sample2	5497776	914899	15	4582862	1384844	3198018	30.2	
D8-Sample3	4710977	789877	19	3921081	1214370	2706711	31.0	
D10d-Sample1	9786341	972550	49	8813742	1410369	7403373	16.0	
D10d-Sample2	5046272	887863	13	4158396	1358339	2800057	32.7	
D10d-Sample3	6176969	1146552	21	5030396	1716458	3313938	34.1	
R3-Sample1	2269038	431391	3	1837644	687580	1150064	37.4	
R3-Sample2	1966504	340563	6	1625935	533487	1092448	32.8	
R3-Sample3	4409262	614660	6	3794596	1188580	2606016	31.3	
R24-Sample1	3580658	819575	8	2761075	1228917	1532158	44.5	
R24-Sample2	5913507	1330397	16	4583094	2078353	2504741	45.3	
R24-Sample3	5071881	1122334	14	3949533	1752922	2196611	44.4	

2.2.2 発現変動遺伝子解析

得られたリード数は tags per million (TPM) により正規化した。この正規化した各サンプル毎の各遺伝子の TPM を反復発現変動遺伝子除去戦略 (iterated DEG Eliminated Strategy, iDEGES) 法 [88] を用いて各サンプル間で発現量の正規化を行った。この際に反復処理時における正規化と仮発現変動遺伝子検出には Trimmed Mean of M value (TMM) および edgeR [89] を用い、反復回数は3回とした。得られた正規化リード数に対して edgeR を用いて発現変動に関する p 値を算出した。この p 値から Benjamini-Hochberg (BH) 法 [90] を用いて False Discovery Rate (FDR) を算出し、 $FDR < 0.05$ となる遺伝子を発現変動遺伝子として検出した。各サンプル間における発現変動遺伝子の発現量に関する分布は M-A Plot [91] で可視化した。

2.2.3 Heatmap および階層的クラスタリングによる時系列遺伝子発現パターンの可視化

全てのサンプル間で検出された発現変動遺伝子の TPM を CRAN の heatmap.2 [92] により発現量の可視化および階層的クラスタリングを行なった。階層的クラスタリングに際しては、距離指標をユークリッド距離、連結法は完全連結法を採用した。

2.2.4 ネムリユスリカゲノムに対する GO アノテーションおよび Fisher の正確確率検定に基づくエンリッチメント解析

明らかにされていたネムリユスリカゲノム [19] においては機能注釈である GO のアノテーションが不十分であったため、本研究では Blast2GO [93] を用いて機能アノテーションを行なった。Blast2GO での解析時の設定としては blastx オプションを利用し、E-value の閾値は $1.0E-05$ とした。GO マッピングおよびアノテーションには blastx の結果を元に CutOff および GO level weighting についてはそれぞれ 30 および 5 と設定し解析を実行した。得られた GO アノテーション結果を用いて、各サンプル間における発現変動遺伝子に対してフィッシャーの正確確率検定を用いて頻出する GO を検出した。この際には、フィッシャーの正確確率検定により得られた p 値を BH 法を用いて補正し、補正後の $FDR < 0.05$ となる GO を検出した。これら検出された GO について、各サンプル間毎に REVIGO (REduced and VISualize Gene Ontology, [94]) を用いて類似した GO の集約を行なった。この際には GO 間の意味論的距離算出には Rel 法 [95] を利用し、距離が 0.7 未満の GO は集約した。これによって集約された GO を各サンプル間の発現変動遺伝子において頻出した GO として本研究では考慮した。

2.2.5 タンパク質ドメイン構造解析

発現変動遺伝子のうちで GO 解析の結果から注目すべきことが示唆された遺伝子に関しては、遺伝子の翻訳後のアミノ酸配列を元に配列類似性を持つ Conserved Domain Database [96] に登録されているタンパク質との間のドメイン領域の類似性を検証した。この検証には multiple sequence alignment viewer (MSAViewer, <https://www.ncbi.nlm.nih.gov/projects/msaviewer/>, [97]) を用いて、ドメイン領域におけるアライメントが形成されているかを確認することで行なった。

2.2.6 Pv11 細胞, ネムリユスリカ虫体および乾燥耐性を持たないヤモンユスリカの比較解析

発現変動遺伝子のうちで GO 解析の結果から注目すべきことが示唆された遺伝子について、Pv11 細胞, ネムリユスリカおよび乾燥耐性を持たないヤモンユスリカの虫体における発現変動と比較した。この際に、ネムリユスリカおよびヤモンユスリカの遺伝子発現量については先行研究で定量された RNA-seq データを用いた [19]。発現変動の有無については同先行研究で用いられた基準である、サンプル間の Fold Change が 3 倍以上かつ RPKM (reads per kilobase of exon per million mapped sequence reads) が 10 以上の遺伝子を発現変動したと判断した。

2.2.7 RT qPCR

Pv11 細胞を IPL-41, 10% FBS を加え、低湿度条件 (RH<10%) に設定した乾燥器に入れ 10 日乾燥させた後に、IPL-41, 10% FBS を再び加え 5 分後 (D10d), 3 時間後 (R3) に細胞を超音波破碎し生物学的反復数 3 でサンプルを調製した。その後 Relia Prep RNA Tissue Miniprep System (Promega, Madison, Wisconsin, USA) を用いて Total RNA を抽出した。得られた RNA を Quantiflour RNA system (Promega) を用いて ssDNA mode で定量を行った。得られた RNA 濃度を元に 50 ng の RNA を含むように各サンプルの希釈を Transcriptor First Strand cDNA Synthesis Kit (Roche Applied Science, Mannheim, ドイツ) に付属する Water, PCR (Vial 7) を用いて行った。得られた各サンプルを Transcriptor First Strand cDNA Synthesis Kit を用い、Anchored Oligo (dT) Random hexamer primer 600 pmol/ μ L (Vial 6) を使用し Thermal Cycler を用いて 50 °C, 60 秒, 85 °C, 5 秒で cDNA 合成を行った直後に氷上にサンプルを静置した。得られた各サンプルの cDNA は Quantiflour cDNA system (Promega) を用いて ssDNA mode で定量を行った。得られた cDNA 濃度を元に Total cDNA が各サンプルで 10 ng になるように MiliQ を用いて希釈を行い SYBR Premix Ex Taq (TaKaRa, 滋賀, 日本) を用いてプライマーの最終濃度が Forward 及び Reverse 共に 0.2 μ M 及び調整した cDNA, SYBR Premix Ex Taq (Tli RNaseH Plus) (2 \times) が 1 \times , Total 20 μ L になるように MiliQ を用いてサンプル調整を行った。

2 Pv11 細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の推定2.3. 結果

本研究では発現変動遺伝子解析および GO 解析で検証の対象となった Pv.07646 に対する発現量の定量を行い、Forward および Reverse のプライマーはそれぞれ、5'-TGAGACAAGACGAGCCAGATG-3' および 5'-CCAAATGAGGAGCGGAATG-3' を用いた。得られた各サンプルを用いて、95°C, 30 秒で調整後、95°C, 5 秒、62°C, 30 秒を 41 サイクル qPCR を実行後、Melt Curve を得た。Control Reaction として no template control を行った。得られた C_t 値を元に D10d に対する R3 の Fold Change を $2^{C_t(D10d) - C_t(R3)}$ により得た。

2.2.8 CAGE-seq データの可用性

本研究で用いた CAGE-seq データは DNA Data Bank of Japan Sequence Read Archive に Accession Number: DRA007433 として登録した (<https://ddbj.nig.ac.jp/DRAsearch/submission?acc=DRA007433>)。

2.3 結果

2.3.1 CAGE-seq データに基づく遺伝子発現量解析結果

Pv11 細胞における乾燥耐性に潜在的に関与する可能性のある遺伝子の検出を行うために、トレハロース処理 0 時間 (T0), Pv11 細胞が乾燥耐性を獲得するために必要なトレハロース処理時間である 48 時間 (T48), その後の乾燥処理後における応答を見るために 8 時間 (D8), 十分な乾燥後の状態を見るために 10 日間 (D10d), 再水和直後の応答を見るために 3 時間 (R3), そして十分に細胞内に水が供給されたと考えられる 24 時間後 (R24) の各サンプル間における発現変動遺伝子解析を行なった (表 2.2)。また、各サンプル間における発現変動遺伝子の M-A Plot の結果から、全てのサンプル間で発現変動遺伝子以外における M 値の中央値が 0 に近い値を取っていることが確認され、M 値がサンプル間の発現量の違いを表すことを考慮すると、この結果は発現変動遺伝子以外をよく除外できていることを示している (図 2.1)。また、D8 と D10d という完全な乾燥状態におけるサンプル間においては発現変動遺伝子が検出されなかった。続いて、Pv11 細胞のこれらの処理過程における遺伝子群の発現変動の遷移を確認するために発現変動遺伝子に対するサンプル間の階層的クラスタリングを行なった (図 2.2)。この結果、完全な乾燥状態である D8 および D10d はトレハロース未処理の T0 に対して発現変動遺伝子の発現パターンが最も異なることが明らかとなった。また、再水和処理後のサンプルである R3, R24 においては段階的に T0 に発現パターンが近づく傾向があった。

2 Pv11 細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の推定2.3. 結果

表 2.2: 各サンプル間における発現変動遺伝子の個数

各サンプルは生物学的反復数 3 で取得され, BH 法による $FDR < 0.05$ の遺伝子を発現変動遺伝子として検出した. サンプル名と取得された条件については 2.2.1 小節を参照.

	T0	T48	D8	D10d	R3
T48	384				
D8	813	37			
D10d	974	44	0		
R3	249	26	25	14	
R24	327	471	704	784	72

2 Pv11 細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の推定2.3. 結果

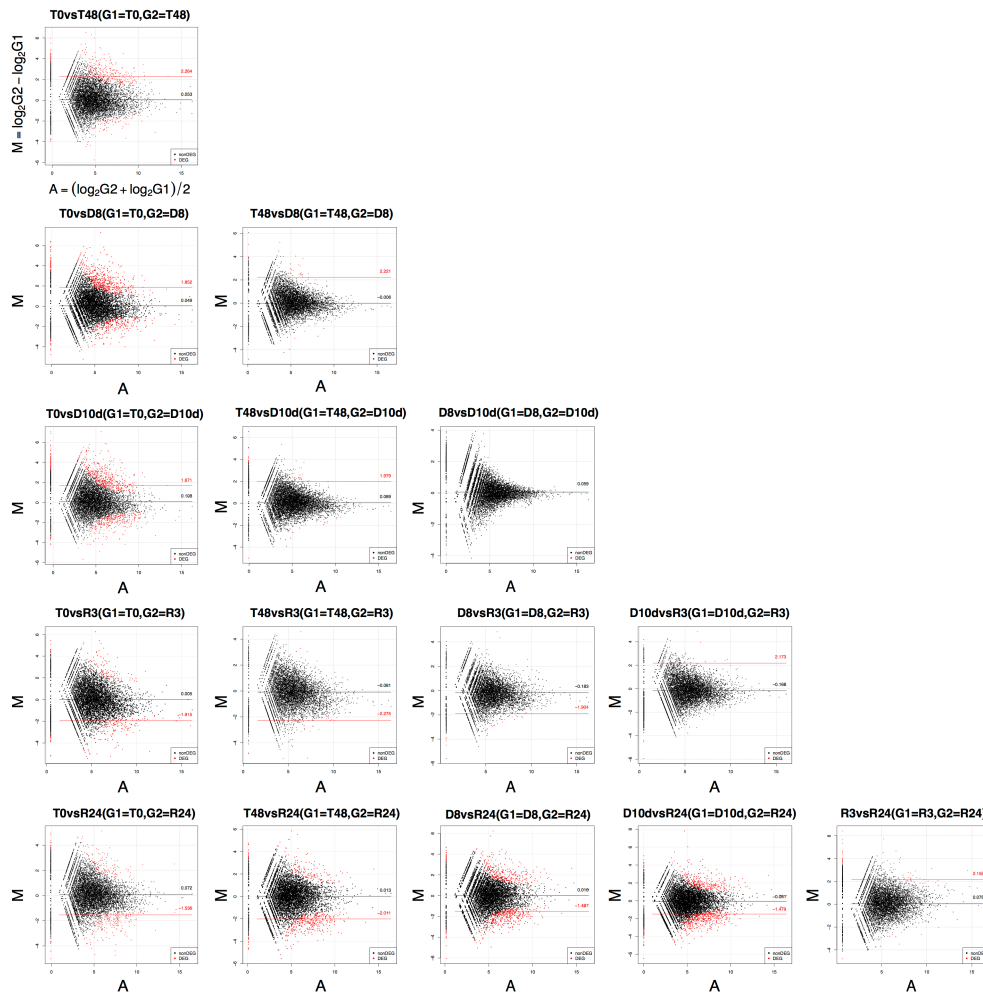


図 2.1: 各サンプル間における発現変動遺伝子に対する M-A Plot

M-A Plot における M 値はサンプル間における遺伝子の log 比 (サンプル G2 に対するサンプル G1 の log 比, $\log_2(G2) - \log_2(G1)$) を表し, A 値は両サンプルの平均値を表し ($(\log_2(G2) + \log_2(G1))/2$), 各図内のプロットは一つの遺伝子に関する M および A 値を示す. 各図内の黒および赤線はそれぞれ, 発現変動遺伝子以外および発現変動遺伝子に対する M 値の中央値を示す. Reprinted from [98]. ©2018 Yamada, Suetsugu, Deviatiiarov, et al., <http://creativecommons.org/licenses/by/4.0/>

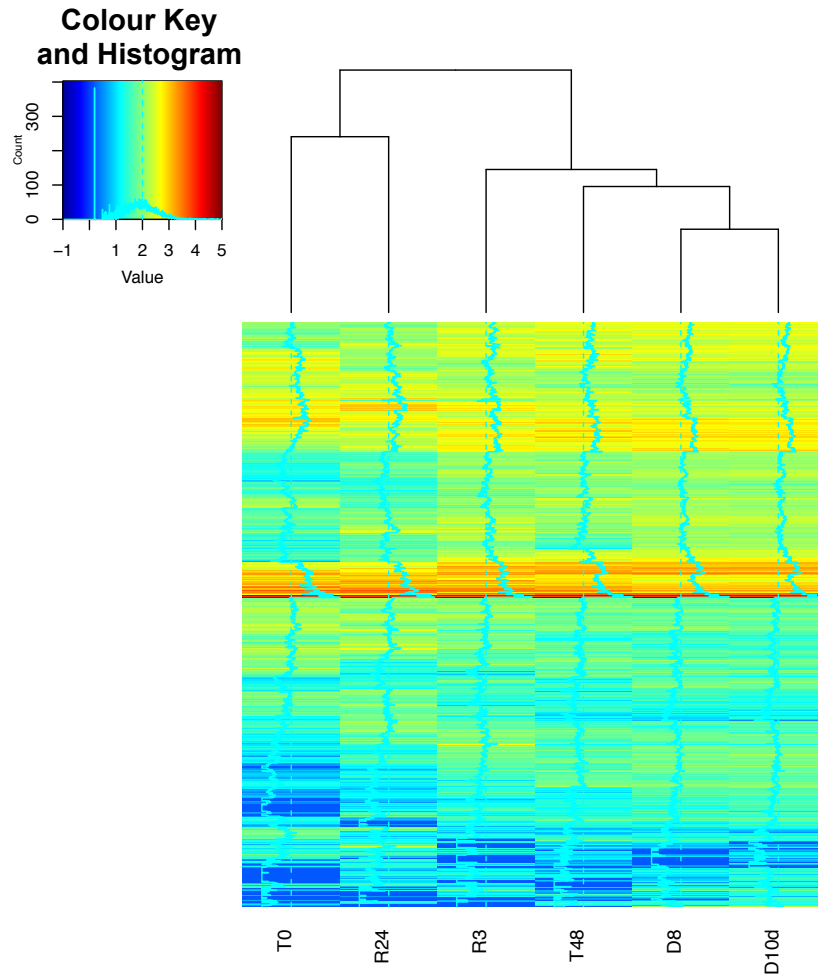


図 2.2: 各サンプルにおける遺伝子発現量 (TPM) に基づく階層的クラスタリング結果

距離指標にはユークリッド距離, 連結法には完全連結法を用いた. Colour Key and Histogram においては, 横軸に $\log_{10}(TPM + (0 \text{ 以外で最小の } TPM \text{ の値}))$, 縦軸にその値を持つ発現変動遺伝子の個数を表した. 図内に存在する遺伝子はいずれかのサンプル間で発現変動遺伝子として検出された. Reprinted from [98]. ©2018 Yamada, Suetsugu, Deviatiiarov, et al., <http://creativecommons.org/licenses/by/4.0/>

2.3.2 発現変動遺伝子に対する GO 解析結果

続いて、Pv11 細胞の乾燥耐性獲得期、無代謝期および再水和期において働く機構がどのような機能を持った遺伝子によって担保されているかを明らかにするために GO 解析を行なった。乾燥耐性獲得期に関しては、T0 と T48, T0 と D8, T0 と D10d, T48 と D8, T48 と D10d 間における GO 解析の結果に注目した。

まず、T0 と T48 間においては 299 個、85 個の遺伝子がそれぞれ有意な発現上昇および発現減少を示した。これらの発現変動遺伝子に対しては頻出 GO は検出されなかった ($FDR > 0.05$)。そこで、このサンプル間においては GO の頻度解析を行ない、最頻出および 2 番目に頻出する GO はそれぞれ、Biological Process に関して GO:0008152 (metabolic process) および GO:0055114 (oxidation-reduction process), Molecular Function に関して GO:0005524 (ATP-binding) および GO:0008270 (zinc ion binding), そして Cellular Component に関して GO:0005634 (nucleus) および GO:0016021 (integral component of membrane) であった。代謝および抗酸化は乾燥耐性獲得期において個体において重要な反応であることを考慮すると (1.2.1.1, 1.2.1.3 小々節参照), これらの GO がアノテーションされた遺伝子は乾燥耐性機構における乾燥耐性獲得に重要な役割を担う可能性が高い。

T0 と D8 間においては 540 個、273 個の遺伝子がそれぞれ有意な発現上昇および発現減少を示した。これら合計 813 個の遺伝子に対する GO 解析の結果、これらの遺伝子に最も顕著に頻出していた 3 つの GO として GO:0005484 (SNAP receptor activity), GO:0004719 (protein-L-isoaspartate (D-aspartate) O-methyltransferase activity) および GO:0006479 (protein methylation) が検出された。この中でも GO:0004719 および GO:0006479 がアノテーションされた遺伝子は PIMT をコードしており、ほぼ全てが D8 において T0 に対して顕著な発現上昇を示した。PIMT はネムリユスリカの虫体においても乾燥時に発現上昇が確認されていることから (1.2.1.3 小々節参照), Pv11 細胞の乾燥耐性獲得期間においても PIMT が重要な役割を担う可能性が明らかになった。

T0 と D10d 間においては 624 個、350 個の遺伝子がそれぞれ有意な発現上昇および発現減少を示した。これら合計 974 個の遺伝子に対する GO 解析の結果、GO:0006412 (translation), GO:0003735 (structural constituent of ribosome) および GO:0019843 (rRNA binding) が最も顕著に頻出した 3 つの GO として検出された。これらの GO がアノテーションされた遺伝子の多くはリボソームタンパク質をコードしており、T0 に対して D10d において顕著に発現減少を示すことが確認された。リボソームタンパク質は遺伝子の翻訳において必要不可欠なタンパク質であり、これをコードする遺伝子の発現量が D10d 時に顕著に発現減少するという結果は、Pv11 細胞が完全な乾燥状態に入るとタンパク質の翻訳を抑えた状態に遷移することを示していると言える。

T48 と D8 間においては 27 個、10 個の遺伝子がそれぞれ顕著な発現上昇および発現減少を示した。これら合計 37 個の遺伝子に対する GO 解析の結果、GO:0005212

2 Pv11 細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の推定2.3. 結果

(structural constituent of eye lens) および GO:0006376 (mRNA splice site selection) の二つが最も低い FDR を示す GO として検出された。この中で特に GO:0005212 がアノテーションされた遺伝子は熱ショックタンパク質 (Heat Shock Protein, HSP) をコードする遺伝子であり、D8 において顕著な発現上昇が確認された。熱ショックタンパク質がストレス誘導障害を緩和する遺伝子であることを考慮すると、Pv11 細胞もこの遺伝子を乾燥耐性機構に組み込むことで、生体にとっての障害を抑える機構を持つ可能性が示唆された。

T48 と D10d 間においてはそれぞれ 32 個、12 個の遺伝子が顕著な発現上昇および発現減少を示した。これら合計 44 個の遺伝子に対する GO 解析の結果、GO:0005212 (structural constituent of eye lens) が最も低い FDR を示す GO として検出された。この GO がアノテーションされた T48 と D10d 間の発現変動遺伝子も T48 と D8 間において得られた発現変動遺伝子と同じ熱ショックタンパク質であることが確認された。

これらの解析結果から、Pv11 細胞の乾燥耐性獲得期においては PIMT や熱ショックタンパク質などのタンパク質に対する障害、ひいては生体にとっての障害となる要因を修復する遺伝子が高発現し、無代謝状態への移行につれてタンパク質を翻訳するために必要不可欠な遺伝子の発現が徐々に抑えられることが明らかとなった。

続いて、完全な乾燥状態であり無代謝期として考えられる D8 と D10 間においては前述した通り発現変動遺伝子は一つも検出されず、GO 解析も行えなかった。このことは顕著な発現上昇、発現減少を示す遺伝子がこの期間になかったということを示しており、無代謝期においては遺伝子制御に基づく転写が行われておらず、さらに乾燥耐性獲得期において準備された転写物が保持されていることを示している。

続いて、再水和期に関しては再水和直後の D10d と R3 間の発現変動遺伝子に着目した。この処理区間では 7 個の遺伝子が有意な発現上昇を示し、7 個の遺伝子が有意な発現減少を示した。これらの合計 14 個の遺伝子に対する GO 解析の結果、8 つの GO が最も低い FDR を示す GO として検出された。これらの GO がアノテーションされた遺伝子のうち Pv.07646 のみ、R3 特異的な発現上昇を示すことが明らかとなった (図 2.3 a, D10d および R24 に対し、35 および 15.5 倍の TPM)。この遺伝子については CAGE-seq による定量結果で分散が大きかったため、RT-qPCR による定量を行い、その結果も D10d に対する R3 における顕著な発現上昇が確認された (図 2.3 b, Welch の t 検定, p 値 < 0.05)。

2 Pv11 細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の推定2.3. 結果

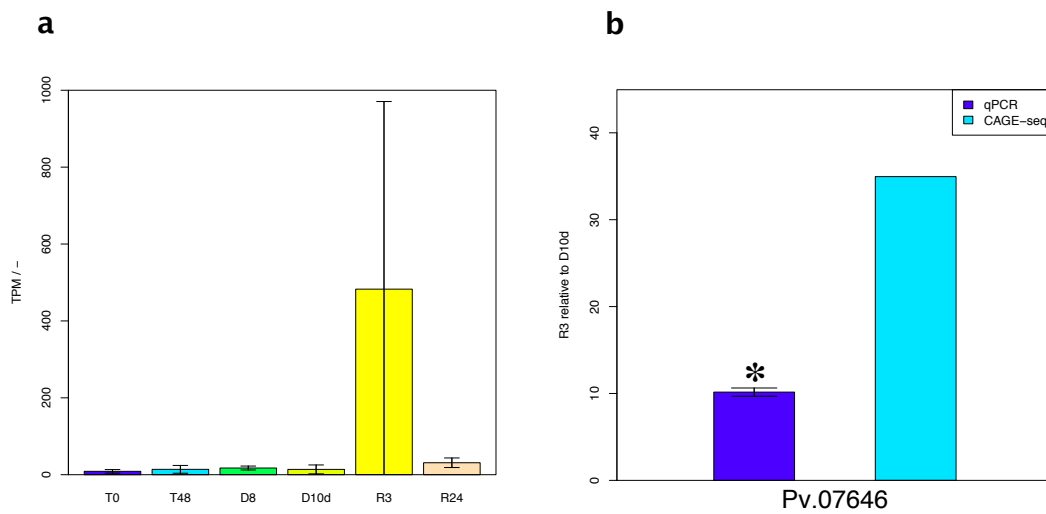


図 2.3: CAGE-seq および RT qPCR によって定量された Pv.07646 の発現量

(a) CAGE-seq で定量された Pv.07646 の発現量 (TPM). 各系列はサンプルを表しており, 縦軸は TPM を表す. 各データは生物学的反復数 3 回の実験の平均 \pm 標準偏差を示す. (b) Pv.07646 の D10d に対する R3 時の CAGE-seq および RT qPCR によって定量された発現量比. *有意な発現変動 (Welch の t 検定 (p 値 < 0.05)). Reprinted from [98].

©2018 Yamada, Suetsugu, Deviatiiarov, et al.,
<http://creativecommons.org/licenses/by/4.0/>

2.3.3 乾燥耐性獲得期に対する解析結果

乾燥耐性獲得期における機構に関与する遺伝子のより深い知見を得るために、T0とT48間で検出されたGO:0008152 (metabolic process; 32個の発現変動遺伝子にアノテーション) およびGO:0055114 (oxidation-reduction process; 26個の発現変動遺伝子にアノテーション) がアノテーションされた遺伝子に注目した。これらの遺伝子をより詳細にグループ化するために、発現変動のパターンに基づいて階層的クラスタリングを行なった。まずGO:0008152がアノテーションされた遺伝子群は4つのクラスタに分類された(図2.4 a)。これらのクラスタにおける全体的な発現変動の傾向を確認すると、トレハロース処理直後に発現上昇を示したこれら遺伝子群は1) 乾燥後にも発現を一定に保つ遺伝子群(クラスタ1)、2) 乾燥後に発現を減少させる遺伝子群(クラスタ2, 3, 4)に大別された。また、GO:0055114がアノテーションされた遺伝子群は3つのクラスタに分類された(図2.4 b)。この結果、これらのクラスタに分けられた遺伝子群は乾燥後に発現上昇(クラスタ3)、発現量を一定に保つ(クラスタ1)、発現減少をするクラスタ(クラスタ2)に大別された。このように乾燥耐性獲得期において重要な役割を担う可能性のある遺伝子群において、同一のGOがアノテーションされた場合であってもトレハロース処理、乾燥、再水和によって異なる発現パターンを示すことが明らかになった。このことは、同一の機能情報を示すGOがアノテーションされた遺伝子であっても異なるタイミングで異なる役割を持つことで、Pv11細胞が乾燥耐性を獲得する上で働く可能性が示された。

この結果を受けて、さらに乾燥耐性獲得期における詳細な機構の理解に向けて、GO:0008152およびGO:0055114がアノテーションされた遺伝子群の機能を配列相同性解析に基づき詳細に検証した。まずGO:0008152がアノテーションされた遺伝子の機能に着目する。Pv.03555はnonsense-mediated mRNA decay (NMD) 遺伝子との間で配列相同性が確認された。NMD遺伝子は不当なストップコドンを持つ未成熟RNAの分解を誘導する遺伝子として報告されている[99]。次にPv.06995はglutathione S-transferase (GST) との間で配列相同性が確認された。GSTは脂質膜が過酸化された場合に、これを分解して無毒化する役割がある[100]。これらの結果から、Pv11細胞の乾燥耐性獲得期においては転写物およびリン脂質膜における障害を修復する遺伝子が重要な役割を担う可能性が示された。GSTは抗酸化物質として知られており、シロイヌナズナに対する干ばつストレス後に高発現する[101]。次にPv.11397はUDP-glucuronosyltransferase (UGT) 遺伝子と配列相同性が確認された。UGTはビリルビン、フェノール、アミンやカルボン酸による毒性を示すグルクロニド化による障害を修復する遺伝子として報告されている[102]。次にPv.04950はCarboxylesterase 遺伝子との間で配列相同性が確認された。Carboxylesteraseはエステルやアミドなどの異物の加水分解を誘導する酵素をコードする遺伝子である[103]。これらの結果から、Pv11細胞の乾燥耐性獲得期においては生体にとって障害となる化合物による影響を除去する遺伝子が重要

2 Pv11 細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の推定2.3. 結果

な役割を担う可能性が示された。

次に Pv.08737 および Pv.12163 はそれぞれ ATP-dependent RNA helicase および DEAD box ATP-dependent RNA helicase との間で配列相同性が確認された。さらにこれらの遺伝子は NMD と配列相同性が見られた Pv.08737 と同じクラスター 1 に属する遺伝子であり、このクラスター内の遺伝子は乾燥時に発現が一定になることが確認されていた。これら 3 つの遺伝子はいずれも転写物の分解に関わる遺伝子であり [104]、乾燥耐性獲得時に発現上昇し、乾燥時にこれら遺伝子の発現が一定で留められていることとあわせて考えると、乾燥時に生じる様々な影響による mRNA への障害にいち早く対応して未成熟な mRNA を速やかに分解する機構がある可能性を示している。

続いて GO:0055114 がアノテーションされた遺伝子の機能に着目する。この GO がアノテーションされた Pv.00404, Pv.06474, Pv.07504 および Pv.07510 は、すでに先行研究で PvTrx1-1, PvTrx3, PvTrx7 および PvTrx10 というチオレドキシンをコードする遺伝子として同定されており、ネムリユスリカ虫体においても乾燥に伴い発現上昇する [19]。この結果は Pv11 細胞においても個体と同様に抗酸化という機構を乾燥耐性獲得時の一要因として利用することを示しており、個体と細胞における乾燥耐性機構の共通性を支持する。

これらをまとめると、トレハロース処理により発現を有意に上昇した遺伝子群は、1) 種々の転写物や脂質膜への障害を修復するストレス応答遺伝子 (NMD Protein, glutathione S-transferase, ATP-dependent RNA helicase および DEAD box ATP-dependent RNA helicase), 2) 生体障害となる化学物質の無毒化を行う遺伝子 (UDP-glucuronosyltransferase や carboxylesterase) 3) 種々のタンパク質の酸化を防ぐためのチオレドキシ遺伝子である。これらの遺伝子が持つ機能は先行研究から判断すると、種々の生体にとって障害となるものを排除する役割を持った遺伝子である。つまり、トレハロース処理による乾燥耐性の獲得とは、種々の生体障害から生体を保護することにより成り立つ可能性がある。

2 Pv11 細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の推定2.3. 結果

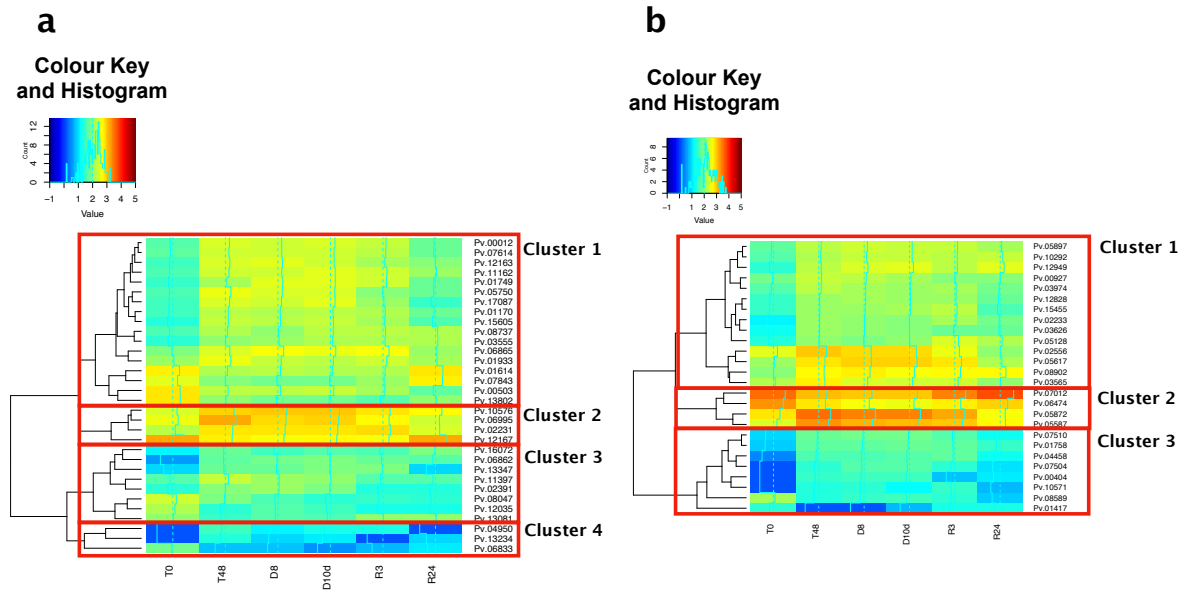


図 2.4: GO:0008152 (metabolic process) および GO:0055114 (oxidation-reduction process) がアノテーションされた遺伝子に対する階層的クラスタリング結果

階層的クラスタリング手法および図の表記法は図 2.2 と同じとした。(a) GO:0008152 (metabolic process) がアノテーションされた遺伝子群に対する各サンプル間の階層的クラスタリング結果。(b) GO:0055114 (oxidation-reduction process) がアノテーションされた遺伝子群に対する各サンプル間の階層的クラスタリング結果。Reprinted from [98].

©2018 Yamada, Suetsugu, Deviatiiarov, et al.,
<http://creativecommons.org/licenses/by/4.0/>

2.3.4 再水和期に対する解析結果

再水和期, 特に再水和直後である R3 特異的に発現上昇を示した Pv.07646 に注目することで再水和期における機構に関する知見を得ることを目指す. Pv.07646 が要因となった GO 解析から得られた GO がマップされることになった由来遺伝子は SNF histone linker PHS ring helicase だった. この遺伝子はヒストンの凝集に関わるリジン残基における翻訳後修飾の一つであるサクシニル化およびマロニル化を解除するデサクシニル化とデマロニル化を誘導する遺伝子である [105]. デサクシニル化とデマロニル化の進行にはヒストンを捉えるために事前に RING ドメインを用いて DNA に結合することと, この反応を SIRT5 領域によって触媒することが必要となる. この反応はヒストンの凝集の有無に関わることから, 近傍遺伝子の転写の有無に関わるため, 遺伝子制御の観点からも非常に重要な遺伝子である可能性がある. そこで, この Pv.07646 がこれらの反応を進行するための必要な条件を備えているかを検討するために, Pv.07646 と SNF histone linker PHS ring helicase のアミノ酸配列に基づきタンパク質ドメイン構造解析を行なった (図 2.5). その結果, Pv.07646 は SNF histone linker PHS ring helicase の RING ドメインにおいてはアライメントが形成された一方で, SIRT5 領域においてはアライメントが形成されないことがわかった. すなわち Pv.07646 はデサクシニル化やデマロニル化を直接進行するための必要な条件を備えていないことが明らかになった. しかしながら, Pv.07646 は再水和特異的に発現することが CAGE-seq および RT-qPCR 実験によって確認されているため, 何らかの再水和特異的な現象に関わる可能性が強く支持されている. そこで, Pv.07646 と配列相同性が最も高い遺伝子を検証したところ, この遺伝子は Rad16 と呼ばれる DNA 修復に関わる遺伝子だった. Rad16 は Rad7 と複合体を形成することで ATP 濃度依存的に DNA 修復を行う [106]. また Rad16 の HepA 領域を失活させると UV 照射を受けた DNA を修復することができなくなることが大腸菌において報告されている [107]. これらの事実から Rad16 が DNA 修復を進行するためには ATP 結合サイトと HepA 領域が必要となる. これらの領域を Pv.07646 も持つかどうかを検証したところ, 両方のサイトにおいて Rad16 との間にアライメントが形成された (図 2.6). このことから, Pv.07646 は Rad16 として DNA 修復に関わることが示唆され, 再水和直後の特異的な発現上昇と統合して, 再水和期の機構として DNA 修復機構に関わる可能性がある.

この再水和期における DNA 修復機構についてさらに深い知見を得るために, DNA 修復機構関連遺伝子の網羅的な解析をした. これまでに先行研究により, DNA 修復機構としては相同組替え (Homologous Recombination, HR), ヌクレオチド除去修復 (Nucleotide Excision Repair, NER), DNA ミスマッチ修復 (Mismatch excision Repair, MMR) および非相同末端結合 (Non-Homologous End Joining, NHEJ) 機構が明らかにされており, それらを構成する遺伝子についても同定されている [108, 109]. そこで, これらの機構を構成する遺伝子と配列相同性を持つ

2 Pv11 細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の推定2.3. 結果

Pv11 細胞のトレハロース処理，乾燥，再水和時における発現変動の有無を検証した (表 2.3). その結果，HR，NER，NHEJ を構成する遺伝子と配列相同性を持つ Pv11 細胞において，これらの処理に応じた有意な発現上昇が確認された．一方で，MMR を構成する遺伝子についてはどの処理においても，一つも発現上昇する遺伝子が確認されなかった．すなわち，Pv11 細胞の乾燥耐性において働く DNA 修復機構としては，HR，NER および NHEJ が働く可能性が示唆された．

2 Pv11 細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の推定2.3. 結果

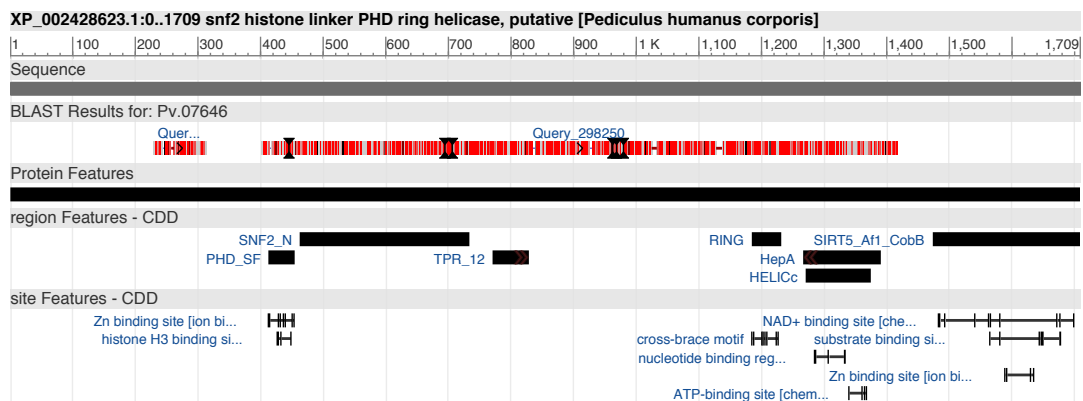


図 2.5: Pv.07646 と SNF2 histone linker PHD-ring helicase のアライメント結果

アライメント結果の可視化には MSASviewer

(<https://www.ncbi.nlm.nih.gov/projects/msaviewer/>) を用いた。 Sequence:

SNF2 histone linker PHD-ring helicase のアミノ酸配列を示す; Blast Results: Pv.07646 と SNF2 histone linker PHD-ring helicase の blastx によるアライメント結果を示す。

region Features: SNF2 histone linker PHD-ring helicase のタンパク質ドメイン領域を示す。 site Features: SNF2 histone linker PHD-ring helicase のタンパク質サイト領域を示す。 BLAST Results: 暗赤色の水平線はギャップ、黒色の垂直線および矢印は挿入、灰色のバーはマッチ領域、赤色のバーはミスマッチ領域を表す。本研究ではタンパク質のドメイン領域とサイト領域と被ってアライメントが形成された場合に、これらの領域が保存されているとした。 Reprinted from [98]. ©2018 Yamada, Suetsugu, Deviatiiarov, et al., <http://creativecommons.org/licenses/by/4.0/>

2 Pv11 細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の推定2.3. 結果

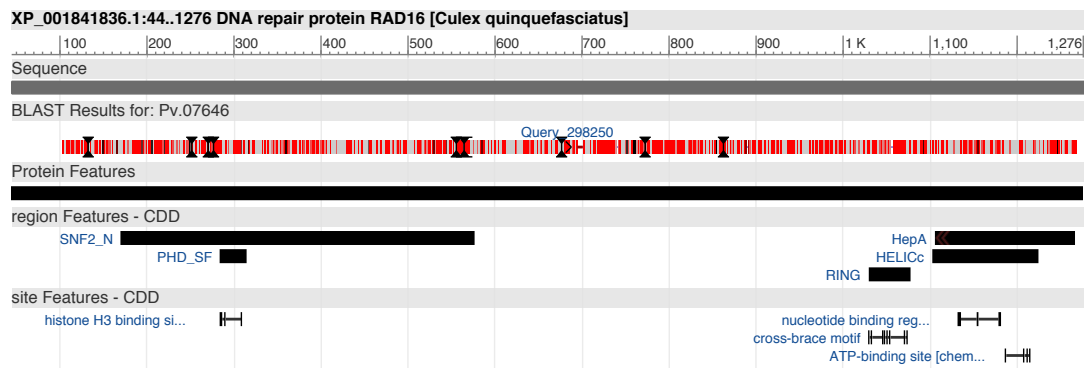


図 2.6: Pv.07646 と DNA 修復遺伝子である RAD16 のアライメント結果

可視化の方法および解釈については図 2.5 と同一とした. Reprinted from [98]. ©2018 Yamada, Suetsugu, Deviatiiarov, et al., <http://creativecommons.org/licenses/by/4.0/>

表 2.3: 種々の DNA 修復機構に関わる遺伝子と配列相同性が確認された遺伝子に関する Pv11 細胞, ネムリユスリカおよびヤモンユスリカにおける発現上昇の有無

Pv11 細胞: トレハロース処理の列は T0 と T48 間の顕著な発現上昇, 乾燥の列は T0 と D8 間, T0 と D10d 間, T48 と D8 間もしくは T48 と D10d 間の顕著な発現上昇, 再水和の列は D10d と R3 間, D10d と R24 間の顕著な発現上昇の有無を表す. ネムリユスリカ: 乾燥の列は D0 と D24 間, D0 と D48 間もしくは D24 と D48 間で顕著な発現上昇, 再水和の列は D48 と R3 間, D48 と R24 間もしくは R3 と R24 間の顕著な発現上昇を表す. ヤモンユスリカ: 乾燥の列は D0 と D24 間の顕著な発現上昇の有無を表す. この際にネムリユスリカおよびヤモンユスリカの結果は [19] を用いた. Pv11 細胞の遺伝子における顕著な発現上昇の有無は M 値が 0 以上かつ $FDR < 0.05$ を基準として判定し, ネムリユスリカおよびヤモンユスリカの遺伝子における顕著な発現上昇の有無は先行研究に慣い, 3 倍以上の RPKM の上昇かつ $RPKM > 10$ を基準として判定した. 各表の内容は顕著な発現上昇が見られた場合には “.” で表示した.

表し, 顕著な発現上昇が見られなかった場合には “.” で表示した.

遺伝子名	Accession Number	ネムリユスリカ遺伝子 ID	E-value	Pv11 細胞		ネムリユスリカ		E-value	ヤモンユスリカ遺伝子 ID	E-value	ヤモンユスリカ
				トレハロース処理	乾燥	再水和	乾燥				
相同組換え (Homologous Recombination)											
RAD51B	NM_002877.5	Pv.04176	1.40E-23	-	T0vsD10d	-	-	-	Pn.00701	4.89E-25	-
RAD50	NM_005732.3	Pv.02239	3.30E-127	-	-	-	-	-	Pn.14177	1.60E-122	-
RAD51	NM_002875.4	Pv.04176	2.88E-150	-	T0vsD10d	-	-	-	Pn.00701	1.33E-153	-
XRCC3	NM_005432.3	Pv.04176	1.30E-16	-	T0vsD10d	-	-	-	Pn.13575	4.00E-22	-
DMC1	NM_007068.3	Pv.04176	1.27E-104	-	T0vsD10d	-	-	-	Pn.00701	9.47E-105	-
ヌクレオチド除去修復 (Nucleotide Excision Repair)											
XPA	NM_000380.3	Pv.10801	1.00E-57	-	-	D10dvsR24	-	-	Pn.08841	5.86E-49	-
ERCC3(XPB)	NM_000122.1	Pv.05115	0.0	-	T0vsD8, T0vsD10d	-	-	-	Pn.12708	0.0	-
GTF2H4	NM_001517.4	Pv.05985	5.29E-147	-	T0vsD10d	-	-	-	Pn.12750	1.07E-148	-
GTF2H2	NM_001515.3	Pv.01131	7.35E-128	-	T0vsD8	-	-	-	Pn.09211	5.95E-130	-
CCNH	NM_001239.3	Pv.01957	2.63E-69	-	-	D10dvsR24	-	-	Pn.07710	3.09E-71	-
非相同末端結合 (Non-Homologous End Joining)											
XRCC6(Ku70)	NM_001469.4	Pv.05333	1.17E-47	-	-	-	-	-	Pn.12095	1.65E-47	D0vsD24
XRCC5(Ku80)	NM_021141.3	Pv.10272	1.05E-41	-	T0vsD8, T0vsD10d	-	-	-	Pn.08610	7.64E-41	-
LIG4	NM_002312.3	Pv.15195	6.24E-62	-	T0vsT48, T0vsD8, T0vsD10d	-	-	-	Pn.05388	1.01E-78	-

2.3.5 Pv11 細胞, ネムリユスリカおよびヤモンユスリカの発現変動遺伝子に関する比較解析結果

これまでに乾燥耐性を持つネムリユスリカ虫体及び乾燥耐性を持たないヤモンユスリカのトランスクリプトームデータに対する比較解析により, ネムリユスリカの乾燥耐性においては LEA, TRX, PIMT, HB, AQP 及びトレハロース産生に関わる種々の酵素が重要であることが示されてきた [19]. このように乾燥耐性に関わる遺伝子を同定するためには, 乾燥耐性を持つネムリユスリカ及び乾燥耐性を持たないヤモンユスリカとの比較解析が有効である. そこで本節では Pv11 細胞における乾燥耐性を明らかにするために, 先行研究で重要とされたこれら遺伝子群, 及び本研究で注目したトレハロース処理, 乾燥及び再水和時に発現変動を示した遺伝子群の発現変動を Pv11 細胞, ネムリユスリカ及びヤモンユスリカの間で比較した. まずネムリユスリカ虫体における乾燥耐性に重要とされた LEA, TRX, PIMT, Hb, AQP 及びトレハロース産生に関わる遺伝子の Pv11 細胞における発現変動を解析した (表 2.4). LEA, TRX, PIMT の一部及びトレハロースの分解に関わる TREH は Pv11 細胞で有意な発現変動を示した一方で, HB, AQP 及びトレハロース合成に関わる TPS はトレハロース処理, 乾燥及び再水和時にいずれも発現変動を示さなかった. また, LEA, TRX, PIMT に関しても, ネムリユスリカ虫体で発現変動を示した遺伝子数よりも Pv11 細胞で発現変動を示した遺伝子数の方がはるかに少なかった. 続いて, 本研究で注目した乾燥耐性獲得時に関わると推定された, Pv.03555 (NMD), Pv.06995 (GST), Pv.11397 (UGT), Pv.04950 (carboxylesterase), Pv.08737 (ATP-dependent RNA helicase) 及び Pv.12163 (DEAD box ATP-dependent RNA helicase) のネムリユスリカ及びヤモンユスリカでの発現変動を解析した (表 2.5). これら遺伝子と配列相同性を持つヤモンユスリカの遺伝子は存在しなかった (blastn E-value < 0.05). また, ネムリユスリカ虫体においては Pv.03555 (NMD) 及び Pv.04950 (carboxylesterase) は再水和時に有意な発現上昇が確認された一方で, そのほかの遺伝子に関しては有意な発現変動はなかった. 続いて, Pv11 細胞の再水和時に重要である可能性が確認された DNA 修復経路に関わる遺伝子群のネムリユスリカ及びヤモンユスリカでの発現変動を解析した (表 2.3). その結果, ヤモンユスリカでは配列相同性を持つ遺伝子が存在する一方で, 乾燥に伴い発現変動する遺伝子はなかった. 続いてネムリユスリカにおいては Pv11 細胞で認められた HR 及び NER に関わる種々の遺伝子は再水和に伴い発現上昇が確認された一方で, NHEJ に関わる遺伝子の発現上昇はなかった. これらの結果から, Pv11 細胞及びネムリユスリカの乾燥耐性に DNA 修復が関わる可能性が示唆された一方で, それぞれの乾燥耐性に関わる DNA 修復経路は異なることが明らかとなった.

2 Pv11 細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の推定2.3. 結果

表 2.4: 先行研究で示された乾燥耐性関連遺伝子の Pv11 細胞, ネムリユスリカおよびヤモンユスリカにおける発現変動遺伝子数
各生物種におけるトレハロース処理, 乾燥, 再水和の列に関しては表 2.3 と同じ定義とした. 表の内容については, 左側の数字が顕著な発現変動を示した遺伝子数を表し, 右側の数字がファミリー遺伝子の総数を表す.

遺伝子ファミリー名	Pv11 細胞			ネムリユスリカ		ヤモンユスリカ
	トレハロース処理	乾燥	再水和	乾燥	再水和	乾燥
LEA	3/23	5/23	0/23	20/23	0/23	0/0
TRX	3/25	7/25	1/25	20/25	0/25	0/3
PIMT	3/14	6/14	1/14	12/14	1/14	0/1
HB	0/32	0/32	0/32	8/32	23/32	3/25
AQP	0/5	0/5	0/5	2/5	2/5	1/5
Trehalose Synthesis	1/4 (TREH)	1/4 (TREH)	0/4	2/4 (TPS, TREH)	0/4	0/4

表 2.5: 本研究で示唆したPv11 細胞の乾燥耐性に関わる遺伝子に関する Pv11 細胞, ネムリユスリカおよびヤモンユスリカにおける発現上昇の有無
 各生物種におけるトレハロース処理, 乾燥, 再水和の列および表内の内容の表示法に関しては表 2.3 と同じ定義とした. また NA は配列
 相同性が得られなかった遺伝子を表す (blastn E-value < 1.0E-5).

遺伝子名	ネムリユス リカ 遺伝子 ID	Pv11 細胞				ネムリユスリカ		ヤモンユス リカ
		トレハロース 処理	乾燥	再水和	乾燥	再水和		
NMD	Pv.03555	T0vsT48	T0vsD8, T0vsD10d	-	-	D48vsR3	NA	乾燥
GST	Pv.06995	T0vsT48	-	-	-	-	NA	NA
UGT	Pv.11397	T0vsT48	-	-	-	-	NA	NA
Carboxylesterase	Pv.04950	T0vsT48	T0vsD10d	-	-	D48vsR24, R3vsR24	NA	NA
ATP- dependent RNA helicase	Pv.08737	T0vsT48	T0vsD10d	-	-	-	NA	NA
DEAD box ATP- dependent RNA helicase	Pv.12163	T0vsT48	T0vsD8, T0vsD10d	-	-	-	NA	NA

2.4 考察

2.4.1 別種との比較解析から考えられる Pv11 細胞とネムリユスリカの乾燥耐性の類似点及び相違点

本研究ではネムリユスリカ虫体及びヤモンユスリカでの発現変動遺伝子に対する比較解析を行なった。先行研究によりネムリユスリカ虫体の乾燥耐性に重要であると同定された遺伝子群のうち、LEA, TRX, PIMT 及び TREH は Pv11 細胞においても発現変動を示す一方で、AQP, HB 及び TPS の発現変動はなかった。Pv11 細胞は胚由来培養細胞であるが、現在までに生体内におけるどの細胞由来かは明らかにされていない。先行研究によりネムリユスリカの脂肪体がトレハロース合成に関わることが報告されている [110,111]。Pv11 細胞がトレハロース合成を行うために必要な TPS を発現上昇させられないことから、Pv11 細胞は脂肪体を構成する細胞とは異なる細胞である可能性が考えられる。また、このことは Pv11 細胞がトレハロース処理を行わないと乾燥耐性を持つことができないことの原因であると考えられる。また、LEA, TRX, PIMT に関して、Pv11 細胞における発現上昇を示すこれらの遺伝子はネムリユスリカの遺伝子に対して少数であった。ネムリユスリカで発現上昇を示したこれら遺伝子のうち、Pv11 細胞で発現上昇を示さなかった遺伝子に注目すると、その発現は T0 の段階で高い発現を示していた (図 2.7)。Pv11 細胞は通常時には IPL41 培地で培養しているが、この培地は生体内よりも塩濃度は高く、さらに先行研究から LEA4 は塩ストレスに応じて発現が誘導されることが報告されている [7, 110, 112, 113]。これら事実から、Pv11 細胞は T0 の段階で塩ストレスに応答した状態になっているため、その際に必要な種々の遺伝子が発現上昇をしているために、トレハロース処理時に発現上昇を示す遺伝子が少なくなっていると考えられた。

また、再水和時に重要と考えられた DNA 修復機構に関しては、Pv11 細胞及びネムリユスリカの両方で HR 及び NER に関わる遺伝子と配列相同性がある遺伝子の有意な発現上昇が確認された。一方で、NHEJ に関わる遺伝子の有意な発現上昇は Pv11 細胞でしか確認されなかった。NHEJ は二重鎖 DNA が切断された際に、DNA の末端を直接つなぎ合わせることで修復する [114,115]。この修復方法は HR や NER よりも修復速度は速いことが知られている [116]。Pv11 細胞においてのみこの修復経路に関わる遺伝子が有意な発現上昇を示したことを考えると、Pv11 細胞はネムリユスリカに対して DNA 損傷に対して鋭敏に応答を示す可能性がある。

Polypedilum 属以外の乾燥耐性を持つ生物種における乾燥耐性関連遺伝子との比較を行った。先行研究により、線虫 (*Caenorhabditis elegans*) や出芽酵母 (*Saccharomyces cerevisiae*) においては LEA 遺伝子やチオレドキシシン遺伝子を乾燥時に高発現する [25,26,117]。本研究で対象とした Pv11 細胞においても LEA 遺伝子やチオレドキシシン遺伝子はトレハロース処理に応じて顕著な発現上昇が誘導される。このことは、トレハロースに加えて LEA 遺伝子やチオレドキシシン遺伝子も個体レ

2 Pv11 細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の推定2.4. 考察

ベルにとどまらず，細胞単位の乾燥耐性においても保存されて重要な役割を担う可能性がある．

2 Pv11 細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の推定2.4. 考察

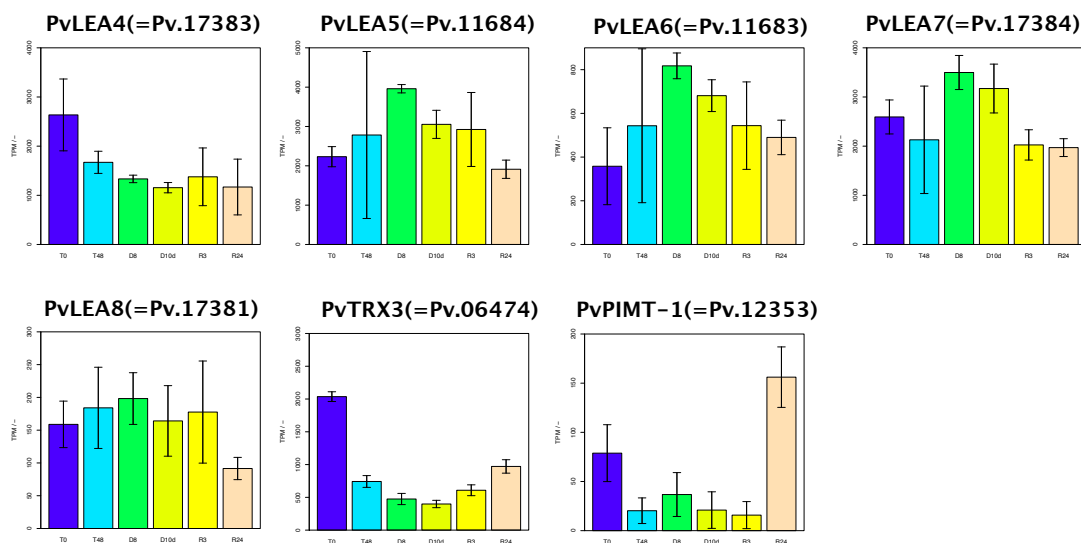


図 2.7: トレハロース処理時に顕著な発現上昇が確認されなかった乾燥耐性関連遺伝子群の時系列発現

これら LEA 遺伝子, TRX 遺伝子および PIMT 遺伝子は, トレハロース処理前から高い発現量を示しており, このことからトレハロース処理後の顕著な発現上昇が確認できなかった. 先行研究ではネムリユスリカの乾燥耐性関連遺伝子は塩ストレスに応じて発現上昇を示すことが確認されており [7, 110, 112, 113], Pv11 細胞も塩濃度が高い IPL-41 培地で培養されている段階で既に乾燥耐性関連遺伝子の一部が発現上昇を示していると考えられた. 平均 \pm 標準偏差で値を示す. Reprinted from [98]. ©2018 Yamada, Suetsugu, Deviatiiarov, et al., <http://creativecommons.org/licenses/by/4.0/>

2.4.2 本研究から得られた Pv11 細胞の乾燥耐性機構と構成する遺伝子群

先行研究ではトレハロースを十分に含まない培地 (443 mM) にて Pv11 細胞を処理すると、再水和後の生存率は 0% であり、トレハロースの濃度を上げる (600 mM) ことで生存率が 16% になることが報告されている [18]. このことから、十分な濃度によるトレハロース処理は、Pv11 細胞に対する乾燥耐性を持たせるために必要である。トレハロース自体は水置換仮説において述べられているように生体物質の防御を行うことで乾燥による死を回避するために必要だと考えられてきた [1, 118]. 一方、本研究の解析結果では、トレハロース処理前とトレハロース処理後 48 時間の間に 384 個の DEG が検出された。このことはトレハロース処理に応答して発現制御を受ける遺伝子が多数存在していることを示している。すなわち、トレハロースは単に生体物質の保護を行なっているだけではなく、乾燥耐性の獲得に必要な種々の遺伝子の発現制御を担う役割を持っていることが新規に示唆された。

また、ネムリユスリカの幼虫においては乾燥直後からトレハロースの蓄積が始まる一方で、この蓄積の間には個体内の水分量はある程度一定に保たれている (2 日間の乾燥では 32 時間、7 日間の乾燥では 6 日より前まで) [110]. この間は個体内の水にトレハロースが存在する状態が続いていると考えられる。すなわち Pv11 細胞におけるトレハロース処理とは、ネムリユスリカにおけるトレハロースの蓄積と水が一定に保たれている期間に該当していると考えられる。すなわちネムリユスリカの幼虫においても、トレハロースによる種々の乾燥寛容性獲得に関わる遺伝子群がこのタイミングで発現を促されている可能性がある。

最後に本研究により得られた Pv11 細胞の乾燥耐性機構を構成する遺伝子群についてまとめる。本研究により、乾燥耐性獲得期においては、これまでの個体の乾燥耐性にも重要であることが知られている LEA 遺伝子、TRX 遺伝子、PIMT 遺伝子が Pv11 細胞においても顕著な発現上昇が確認されたが、発現変動を示す遺伝子はネムリユスリカの個体と比べて少なかった。このことは、個体での乾燥耐性における生体障害緩和機構の冗長性が存在しない、より単純な生体障害緩和機構により Pv11 細胞の乾燥耐性機構が成り立っている可能性を示す。また個体ではまだ乾燥耐性への関与が確認されていない新規候補遺伝子として、種々の転写物や脂質膜への障害を修復するストレス応答遺伝子 (NMD Protein や glutathione S-transferase), 2) 生体障害となる化学物質の無毒化を行う遺伝子 (UDP-glucuronosyltransferase や carboxylesterase) が挙げられ、これらの多くは Pv11 細胞特異的に発現上昇が確認された。このことは Pv11 細胞が乾燥に対して、より幅広い障害ストレスに対する生体障害緩和機構を Pv11 細胞が備えていることが考えられた。これらの考察を総括すると、Pv11 細胞の乾燥耐性における生体障害緩和機構は、個体よりも浅く広い範囲の障害ストレスに対する対策を持った機構である可能性がある。

第3章 Pv11細胞の乾燥耐性を支える マスター転写因子と転写制御 機構の推定

3.1 背景

生命活動に水は必須である。水が失われる乾燥ストレスは多くの生物にとって致死的な要因の一つとなる。特に陸生の生物はこの乾燥ストレスに常に晒されることから、これによる死を回避するために進化の過程で乾燥ストレスを寛容化できる分子メカニズムを獲得してきた。特に乾燥ストレスを寛容化するメカニズムの一つである無代謝状態 (Anhydrobiosis) への移行は99%以上の生体内の水を失ったとしても、水を再度与えられることで元の生活環に戻る仕組みである。これまでにワムシ、クマムシ、線虫や植物、そしてネムリユスリカの幼虫など様々な生物がこの仕組みを持つことで乾燥ストレスによる死を寛容することが知られている (1.2節参照)。さらに近年樹立されたネムリユスリカの胚由来培養細胞である Pv11細胞は、事前に高濃度トレハロースを含む培地処理することで、乾燥ストレスを与えても再び水を与えることで再度細胞分裂を再開し、この乾燥耐性が細胞レベルで達成される仕組みであることが示された [7,18]。

ネムリユスリカの乾燥耐性は、まず乾燥ストレス受容後のトレハロース産生から始まる。水置換仮説に基づくと、乾燥後に生体膜と水との水素結合による相互作用を生体にとって無害である無還元糖の一種であるトレハロースが代替することで細胞膜の構造が安定化され細胞死を防ぐ働きを持つと考えられている。このトレハロース産生後に、生体にとって障害となる活性酸素 (ROS) を除去する抗酸化遺伝子であるチオレドキシシンや、乾燥時に起きるタンパク質の変性を抑える LEA 遺伝子や PIMT 遺伝子の発現が向上する (1.2.1小節参照)。Pv11細胞は自発的なトレハロース産生を行えないため乾燥耐性を獲得させるためにトレハロースを豊富に含む培地で処理する必要があるが、興味深いことにこのトレハロース処理を行うことでチオレドキシシンや LEA, PIMT の高発現が誘導される (2.3.5小節参照)。また、こうした生体にとって有害となる要素の排除に関わる遺伝子の高発現に加えて、乾燥時に欠乏する酸素の保持のためのヘモグロビン遺伝子の高発現や水が欠乏することによる細胞膜上の浸透圧調節のためのアクアポリン遺伝子の高発現が乾燥後に見られる。さらに乾燥が進むとリボソーム関連遺伝子の発現低下が確

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.1. 背景

認められ、乾燥時に不要なタンパク質翻訳のエネルギー消費を抑えている可能性がある (2.3.3 小節参照). そして乾燥から 48 時間程度かけて完全な無代謝状態に入ると極端な発現変動する遺伝子はなくなり、各遺伝子の発現は一定の状態で維持されることが Pv11 細胞において確認された. この状態で再び水が与えられると、3~24 時間程度で DNA 修復機構である相同組替え、ヌクレオチド除去修復および非相同末端結合に関与する遺伝子が高発現する (2.3.4 小節参照). これら遺伝子の高発現は、先に示した生体障害除去機構でも抑えることができなかった DNA への障害を回復することで復活の段階を助けていると考えられた.

このように、乾燥してから水を与えられて生き返るまでの過程に関与する遺伝子については、2 章の研究も含めて生理学的な観点からも多くの研究が行われ、その分子メカニズムの多くが解明されてきた [19, 53, 110, 112, 113, 118–120]. 一方で、この分子メカニズムを駆動することができる転写因子、そしてその転写制御システムについてはいまだ不明な点が多い. これに対して一昨年に、P. Mazin らによりこれら乾燥耐性に関与する遺伝子群のエンハンサーサイトに対する網羅的な解析が行われ、熱ショック因子 (Heat Shock Factor) の結合サイトがこれらの領域に豊富に存在していることが示された. さらに彼らは、HSF を Pv11 細胞において knock down し Wild Type との間で乾燥耐性関連遺伝子の発現を比較したところ、LEA やチオレドキシシン、トレハロース産生に関わるトレハロース 6 リン酸化酵素 (TPP) の発現が有意に減少した. これにより HSF が乾燥耐性において必要な遺伝子制御を担っていることが示されたと言える (1.2.1.3 小々節参照). これは一見、乾燥耐性を支配する転写因子とその遺伝子制御の全貌が明らかになったように見えるが、実際には HSF は乾燥耐性を持たないショウジョウバエなどの生物にも存在する. このことから、HSF の存在のみではショウジョウバエなどが乾燥耐性を持たず、ネムリユスリカなどが乾燥耐性を持っている理由は説明できない. つまり、HSF は乾燥耐性において必要な転写因子ではあるが、これだけでは乾燥耐性機構を支えるのに十分とは言えず、乾燥耐性を持つ生物特有の他の転写因子による転写制御が存在することを示している.

この未知な点を明らかにするために、本研究では Pv11 細胞に対するトレハロース処理、再水和時に介在する遺伝子制御ネットワークを明らかにするために、Pv11 細胞に対する当該処理後の 2 章よりもより詳細な時系列 RNA-seq データを解析した. この時系列 RNA-seq データに対して発現変動を示す遺伝子を検出した. これらの発現変動遺伝子のうち転写因子については回帰型転写制御ネットワーク推定手法を用いて転写制御ネットワークを推定し、それ以外の遺伝子についてはまず協調発現を示す遺伝子集団をモジュールとしてまとめた. そして、これらの転写因子とモジュールの間の制御関係を末端遺伝子制御ネットワーク推定手法を用いることで推定し、乾燥耐性制御に関わる遺伝子制御ネットワークを推定した. 本研究では Pv11 細胞に対する乾燥耐性制御に関わる遺伝子制御ネットワークを推定することで、乾燥耐性関連遺伝子群全体を制御する可能性のある転写因子およびこれら遺伝子群の発現の ON/OFF を切り替える制御構造に関する示唆を行なった.

3.2 方法

3.2.1 提供された RNA-seq データの概要及びデータ前処理

本研究では Pv11 細胞に対するトレハロース処理および再水和時における RNA-seq データを農業・食品産業技術総合研究機構の黄川田研究室より提供いただいた。まず提供いただいたデータの概要について説明する。Pv11 細胞は FBS を 10% 加えた IPL-41 培地にて培養し (T0), その後 10% IPL-41 培地と 600 mM のトレハロースを含む混合液で 12 (T12), 24 (T24), 36 (T36) 及び 48 (T48) 時間培養した。さらに, この 48 時間の培養を経たのち細胞はデシケータで 10 日間乾燥した (RH < 10%)。10 日間の乾燥後に細胞に 10% FBS を含む IPL-41 培地を加え (R0), 3 (R3), 12 (R12), 24 (R24), 72 (R72) 時間再水和した。これらのサンプル採取は全て 3 回の生物学的反復実験により行われた。用意したサンプルはその後, TruSeq RNA sample prep kit v2 (Illumina, San Diego, CA, USA) にて Total RNA が抽出し, NucleoSpin RNA kit (Macherey-Nagel, Düren, Germany) にて洗浄した。抽出した RNA の質は Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) にて評価した。また抽出した RNA は Qubit 2.0 fluorometer (Thermo Fisher, Waltham, MA, USA) にて定量した。この RNA から一本鎖 cDNA 合成が行われ Illumina HiSeq1500 (Illumina, San Diego, CA, USA) (HO mode w/ cBot, v4) にてシングルエンドでシーケンスした。ここまでの処理は黄川田研究室にて実施された実験概要となる。得られたリードデータは FastQC にて Quality Check を行った [121]。Quality Check の結果に基づき, FaQCs [122] を用いて 1) 塩基配列内のアダプター配列領域の切り取り, 2) 20 塩基の連続した 'A' を PolyA 構造として切り取り, その結果リード長が 14 塩基以下になったリード及び 'N' が 2 つ以上含まれるリードを除去した。そして Quality Control 後のリードをネムリユスリカゲノム (midgeBase: <http://150.26.71.110/midgebase/>) に対して HISAT2 [123] を用いてマッピングした。このマッピング結果を用いて, 全リード数に対するマッピングされたリード数をマッピング率として各サンプルについて算出した (表 3.1)。得られたマッピング結果と先行研究により再解析が行われたネムリユスリカのゲノム情報 [39] を用いて Bioconductor の QuasR パッケージ [87] に付属する qCount 関数を用いて各遺伝子上のリードをカウントした。

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.2. 方法

表 3.1: 各サンプルにおけるマッピング結果

サンプル名	マッピングされたリード数	マッピングされなかったリード数	マッピング率 (%)
トレハロース処理			
T0_1	8487568	506228	94.3
T0_2	11803186	675649	94.6
T0_3	20835951	1129919	94.9
T12_1	15369827	797712	95.1
T12_2	9602418	505249	95.0
T12_3	14969341	833484	94.7
T24_1	12381278	661013	94.9
T24_2	14396344	739811	95.1
T24_3	21654754	1082941	95.2
T36_1	10926300	547389	95.2
T36_2	10975957	610096	94.7
T36_3	14512336	739135	95.2
T48_1	22602049	1055999	95.5
T48_2	18609332	872098	95.5
T48_3	8794809	408010	95.6
再水和			
R0_1	12068620	887506	93.1
R0_2	11406662	711955	94.1
R0_3	11676711	810103	93.5
R3_1	8056810	5558020	59.2
R3_2	8680054	5349949	61.9
R3_3	7208915	5951408	54.8
R12_1	12569469	2043604	86.0
R12_2	10011663	2667331	79.0
R12_3	12482169	2976002	80.7
R24_1	11676557	1838090	86.4
R24_2	11893539	1532843	88.6
R24_3	12235718	1426672	89.6
R72_1	13576009	900856	93.8
R72_2	12966806	827886	94.0
R72_3	10530947	698720	93.8

3.2.2 主成分分析

各サンプルにおける大局的な遺伝子発現パターンを確認するために、得られたRNA-seqデータに対する主成分分析を行なった。まず各遺伝子のリード数を元に Reads Per Kilobase of exon per Million mapped sequence reads (RPKM) 法を用いて正規化を行った。この正規化後リード数を用いて、Rに標準実装されている `prcomp` 関数を用いて主成分分析を行った。得られた主成分分析の結果における、各サンプルの第一から第三主成分における主成分得点を算出し、`scatterplot3D` パッケージ [124] を用いて三次元プロットした。

3.2.3 発現変動遺伝子解析

元のカウントデータを用いてトレハロース処理、再水和後のサンプルそれぞれについて、`DESeq2` パッケージ [125] の `nbinomLRT` 関数を用いて尤度比検定により発現変動遺伝子解析を行った。尤度比検定においては各遺伝子に対して全時刻における平均の発現量のみで説明されるモデル (Null Model) と各時刻の平均発現量にパラメータを持つモデル (Full Model) との間の尤度を比較することによって以下の仮説検定を行なった。

H_0 : 全時刻における平均発現量は同一 (帰無仮説)

H_1 : 少なくとも一時刻において平均発現量は他の時刻の平均発現量とは異なる (対立仮説)

この際に、片側検定によるBH法 [90] を用いた補正後p値が0.05未満の遺伝子を発現変動遺伝子として検出した。この発現変動遺伝子のうち、先行研究において転写因子のアノテーション時に用いられるGO:0003700 (sequence-specific DNA binding transcription factor activity) がアノテーションされた遺伝子を転写因子として選別した [126]。そして転写因子とそれ以外の遺伝子について発現変動遺伝子の個数をVennDiagramパッケージ [127] の `venn.diagram` 関数を用いてベン図により描画した。

3.2.4 時系列発現に基づく転写制御ネットワーク推定

発現変動遺伝子として検出された転写因子について、トレハロース処理及び再水和時の時系列のRPKMを用いてBTNET [71] により転写制御関係を推定した (1.3.2.2小々節参照)。BTNETでの転写制御関係の推定は、ブースティング法として勾配ブースティング法、ブースティング木の数を100個、タイムラグを1というパラメータ設定で行った。得られたそれぞれの制御関係ごとのスコアについて、制御関係の有無を決定するための閾値は大腸菌や酵母の転写制御ネットワークがスケールフリー性を持つという先行研究の知見を利用し [128]、以下の数式で記述

されるスケールフリー性へのフィッティングを利用して求めた。

$$p(k) = Ck^{-\gamma} \quad (3.1)$$

(3.1) 式の、 k は各転写因子の次数、 γ はスケールフリー性を示す係数、 C は正規化定数を表す。この式は両対数を取ることで以下のような $p(k)$ と k の間について線形な関係の式に変換できる。

$$\log_{10}p(k) = -\gamma\log_{10}k + \log_{10}C \quad (3.2)$$

この数式に基づき、先ほど求めたスコアの値が a 以上の制御関係を採用して得た転写制御ネットワークについて次数とその頻度分布に対して R の `lm` 関数を用いて線形フィッティングした。ここで γ とその F 検定による片側検定の p 値及びフィッティング時の R^2 値を算出した。この作業を a を 0 から 1 までの間、算出されたスコア毎に行い、 γ が負になり、かつ p 値が 0.05 を下回る最小の a を求め閾値とし、このスコア以上の制御関係を採用した。次に、この得られた制御関係について正と負の制御関係を分類するために、制御関係の source と target 遺伝子の時系列発現量を元にタイムラグ 1 の相互相関係数を R に標準実装の `ccf` 関数を用いて算出した。そして、この値が 0 より大きい場合に正、小さい場合に負の制御として解析した。こうして得られたトレハロース処理及び再水和における転写制御ネットワークの片方に少なくとも存在する制御関係を採用し、統合転写制御ネットワークを得た。得られた転写制御ネットワークは `yEd` [129] を用いて転写因子を四角ノード、制御関係をエッジとして可視化した。また、得られた統合転写制御ネットワークから 3 転写因子間の Feed Forward Loop (FFL) 及び Feedback Loop (FBL) 構造を `igraph` パッケージ [130] の `graph.get.subisomorphism.vf2` 関数によって抽出した。この抽出した FFL 構造については、入力から出力への 2 つの制御関係の正負が一致していた場合に Coherent 型、一致していない場合に Incoherent 型として分類した。FBL 構造については、負の制御関係が 0 もしくは 2 つの場合に Positive FBL、1 つもしくは 3 つの場合に Negative FBL として分類した。これら得られた構造の個数がランダムネットワークに対して有意に多いかを検証するために Randomization test を行った。Randomization test では、まずノード数、エッジ数を統合転写制御ネットワークの転写因子、制御関係の数として `igraph` パッケージの `erdos.renyi.game` 関数を用いて Erodös-Renye モデル [131] によりランダムネットワークを生成した。次に、ランダムネットワークの各制御関係の正負の制御の決定については、統合転写制御ネットワークの制御関係の正負の比率を成功確率とした二項分布に従って決定した。これにより得られたランダムネットワークから Coherent, Incoherent 型の FFL 及び Positive, Negative FBL の個数を算出した。この処理を 10000 回繰り返し、統合転写制御ネットワークに存在したそれぞれの構造の個数以上になるネットワークの割合をそれぞれの構造に対する p 値として算出した。

3.2.5 協調発現モジュールの検出と Gene Ontology Enrichment 解析

本研究では転写因子以外の発現変動遺伝子に対して、Weighted Gene Coexpression Network Analysis (WGCNA) [132] による無向共発現ネットワークを解析することで協調発現モジュールを検出した。まず、WGCNA の `pickSoftThreshold` 関数を用いてべき乗値 (*power*) を求めた。この際にはまず、各遺伝子のトレハロース処理及び再水和の時系列 RPKM を元に遺伝子間のピアソン相関係数を算出した。次に全遺伝子間に対してこの操作を行い、ピアソン相関係数行列を得た。そして、この行列の各値を *power* 乗した値から求められる無向協調発現ネットワークを求め、そのネットワークのスケールフリー性に関する R^2 値が 0.75 を超える *power* を求めた。この結果、スケールフリー性を満たす無向協調発現ネットワークを得るための *power* として 7 が採択された。次に、このピアソン相関係数行列の各値を *power* 乗した行列に関して、より頑健に発現パターンの類似性を表すことができる topological overlap measure [133] を前述の行列の各要素について求めた。この行列の各要素の値を 1 から引いた行列を各遺伝子発現パターンの相違性、すなわち距離行列として求めた。この距離行列を用いて、R に標準実装の `hclust` を用いて群平均法により遺伝子間の階層木を得た。ここから協調発現モジュールを検出するために、得られた階層木を WGCNA パッケージの `cutreeDynamic` 関数を用いて分割した。この際に、`cutreeDynamic` 関数では各クラスタに属する最小の遺伝子数を `minClusterSize` として事前に与える必要がある。そこで、本研究ではこの値を 1 ~ 200 まで振り、それぞれにクラスティング指標である Pseudo F [134] を求め、この値が最大となる `minClusterSize`=165 を採用した。これにより 27 個のクラスタが得られ、これを協調発現モジュールとした。得られた各モジュールを構成する遺伝子の機能情報を明らかにするために、モジュールごとに GOstats パッケージ [135] の `GSEAGOHyperGParams` 関数を利用し、*p* 値 < 0.05 になる GO を検出した。

3.2.6 転写因子とモジュール間の制御関係の予測

続いて、転写制御ネットワークと協調発現モジュールの間の制御関係を求めるために、CLOVER [86] によるモジュール内遺伝子のエンハンサー領域における転写因子の結合モチーフの Enrichment 解析を行った。この際に、ネムリユスリカの各転写因子の結合モチーフに関する情報はいまだないため、ショウジョウバエの転写因子の結合モチーフを利用した。まず、CISBP [84], FlyFactorSurvey [83] 及び JASPER2018 [82] から motifDB パッケージ [136] を用いて、それぞれ 134, 291, 140 個の結合モチーフ情報を取得した。次に、これらの結合モチーフを持つ全転写因子のアミノ酸配列を取得し、ネムリユスリカの転写因子のアミノ酸配列との間で双方向に blastp による配列相同性解析を行った。この際に、*E*-value < 1.0e-5 を

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.2. 方法

閾値として用いた。これにより 72 個のネムリユスリカの転写因子とショウジョウバエの転写因子との間で配列相同性が確認された。このショウジョウバエの転写因子の結合モチーフを配列相同性のあるネムリユスリカの転写因子の結合モチーフとして利用した。この結合モチーフと各モジュールの遺伝子の上流領域 100, 500, 1000, 2000, 3000, 4000, 5000, 10000bp を対象として CLOVER による結合モチーフの Enrichment 解析を行った。また、この際にはコントロール配列はネムリユスリカの Scaffold を用い、いずれかの上流領域を対象とした際に片側検定の p 値 < 0.05 となった結合モチーフとモジュールの関係性を抽出した。また、このような結合モチーフの Enrichment が存在する転写因子とモジュールの間に、遺伝子発現の前後関係、すなわち転写因子の発現変動後にモジュール内の遺伝子の発現変動が存在するかどうかを検証するために Granger 因果性検定を行った [77]。Granger 因果性検定を行うために、まずトレハロース処理及び再水和の時系列 RPKM を Z-score に変換した。次に、それぞれについて、転写因子の時系列 Z-score を原因、モジュール内の遺伝子の各時刻ごとの Z-score の平均を結果とした Vector Auto Regressive (VAR) モデルと因果がない場合のモデルを立て、この間で F 検定を行うことで Granger 因果性の有無を検定した。この解析において、VAR モデルは R の vars パッケージ [137] の VAR 関数を用いてモデル化し、タイムラグは VARselect 関数により AUC 最大となるタイムラグを利用した。Granger 因果の有無は、vars パッケージの causality 関数を用いた F 検定から p 値を算出し、BH 法による補正 p 値 < 0.05 となった場合に、その転写因子とモジュールの間には時系列発現に関して Granger 因果性有りとして検出した。

3.2.7 ショウジョウバエ転写制御ネットワークとの比較解析及びネットワーク縮約

得られた統合転写制御ネットワークと modEncode にて公開されているショウジョウバエの転写制御ネットワークの比較を行った (引用 [61] の図 7A で示される転写制御ネットワーク, <http://intermine.modencode.org/release-33/flyRegulatoryNetwork.do> よりダウンロード)。比較に際して、まずショウジョウバエの転写制御ネットワーク内の転写因子のアミノ酸配列に対して、ネムリユスリカの転写制御ネットワーク内の転写因子のアミノ酸配列を元に一方向性の blastp により配列相同性を解析し、 $E\text{-value} < 1.0e10^{-15}$ を配列相同性有りとして検出した。次に、ネムリユスリカの転写因子間で存在した各制御関係について、対象となる制御関係の source と target となる転写因子と配列相同性を持つショウジョウバエの転写因子間において直接もしくは 1 つ以上の転写因子を介した間接的な制御関係がある場合に、両者に共通した制御関係として抽出した。続いて乾燥耐性に関係する転写制御関係を明らかにするために、得られたネットワークの縮約を行った。ネットワークの縮約には、まずショウジョウバエの転写因子と配列相同性を持たなかった転写因子とモジュールとの直接制御のあった転写因子を対象に、1) 自身への制

御が存在しない最上流転写因子, 2) FFL 構造を構成する転写因子, 3) HSF 及び 1), 2) で採択された転写因子と HSF の間を繋ぐ転写因子のみを選別した. 続いてこれらの転写因子間の制御関係においてショウジョウバエでは存在しない制御関係のみを選別した. さらに, この縮約された転写制御ネットワーク内の FFL 構造を一つのブロックとしてまとめ, FFL から HSF への制御関係に存在する転写因子を除き, 最終的な縮約ネットワークを得た.

3.2.8 最上流転写因子と乾燥耐性関連遺伝子群間の制御関係の有無の妥当性検証

本研究では BTNET の推定結果に基づき最上流転写因子の検出と, その転写因子による乾燥耐性関連遺伝子群の制御に関して調べた. しかし, このような推定結果には偽陽性 (本来は制御関係がないはずなのにあるとしてしまう過誤) と偽陰性 (本来は制御関係があるはずなのがないとしてしまう過誤) が含まれるものであり, BTNET においてもこの調和平均である F-measure を最大化しようとする, ベンチマークの問題に対して最も良い場合で偽陽性のなさを示す Precision が 0.359, 偽陰性のなさを示す Recall が 0.307 であることが確認された.

そこで, 今回の推定結果に対して最上流転写因子と最末端にあたる乾燥耐性関連遺伝子群の間の制御関係の有無の検証を, 前述の Precision の値を利用することで試みた. 最上流転写因子が乾燥耐性関連遺伝子群を制御するためには, その間に少なくとも一つのパスが存在する必要がある. いま, 注目した 1 本の制御関係が真である確率を p とした場合に, 任意の 2 ノード間を結ぶパスが少なくとも一つある確率を導く. ただし, 以下では始点ノードを v_s , 終点ノードを v_f , v_s から v_f の間のパスを $\mathbf{P} := \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$, v_s から v_f のパスの本数を n と定義する. またパス $\mathbf{p}_i := \{e_1, e_2, \dots, e_{l_i}\}$ と定義し, e はパスに含まれるエッジ, l_i はパス i に含まれるエッジの本数を表すものとする. まず, 1 本のエッジが真である確率が p の時, 1 本のパス \mathbf{p}_i が真であるためには全てのエッジが真である場合に限るため, 1 本のパス \mathbf{p}_i が真である確率 $p(\mathbf{p}_i, p)$ は以下で与えられる.

$$p(\mathbf{p}_i, p) = \prod_{j=1}^{l_i} p \quad (3.3)$$

$$= p^{l_i} \quad (3.4)$$

続いて, 2 ノード間のパスが少なくとも一つ存在する確率 $p(\mathbf{P}, p)$ は補集合の性質から, 1 から全てのパスが偽だった場合を減算すれば良い. 上式より, 1 本のパスが偽の確率は $1 - p(\mathbf{p}_i, p) = 1 - p^{l_i}$ より, $p(\mathbf{P}, p)$ は以下のように導かれる.

$$p(\mathbf{P}, p) = 1 - \prod_{i=1}^n (1 - p^{l_i}) \quad (3.5)$$

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.3. 結果

(3.5) 式を利用することで、推定した遺伝子制御ネットワークに基づき、最上流転写因子と乾燥耐性関連遺伝子群間の制御関係の有無に関する妥当性検証を行なった。この際に、始点から終点ノードへのパスの検出には python の networkx モジュール [138] における all_simple_path 関数を利用し、始点から終点ノードまでの間のノード数を 20 に制限してパスを検出した。

3.2.9 RNA-seq データの可用性

本研究で用いた RNA-seq データは DNA Data Bank of Japan Sequence Read Archive に Accession Number: DRA008948 として登録した (<https://ddbj.nig.ac.jp/DRAsearch/submission?acc=DRA008948>)。

3.3 結果

3.3.1 トレハロース処理および再水和時系列遺伝子発現量の定量および発現変動遺伝子の検出

Pv11 細胞において完全な乾燥状態では遺伝子発現の顕著な変動が見られていない (2.3.1)。そこで、本研究ではトレハロース処理及び再水和の過程に注目した。まずトレハロース処理および再水和に応答する遺伝子を検出するために、トレハロース処理後 0 (T0), 12 (T12), 24 (T24), 36 (T36), 48 (T48) 時間および、それから乾燥 10 日を経てから再水和後 0 (R0), 3 (R3), 12 (R12), 24 (R24), 72 (R72) 時間後に各 3 反復実験により total RNA が採取され RNA-seq によりデータが取得、提供された。得られた RNA-seq データをネムリユスリカゲノムにマッピングすることで、各サンプルのマッピング率が R3, R12, R24 の処理区を除いて 90% 以上であった (表 3.1)。続いて、これらのマッピング結果及びネムリユスリカのゲノム情報を用いて全遺伝子の発現量を定量し、主成分分析により各処理区間の遺伝子発現パターンを確認した (図 3.1)。この結果から、処理区間の分散に対して処理区内の分散が十分に小さいことが定性的に確認でき、サンプル採取は十分に統制が取れていることを確認した。

これらのサンプルを用いて、トレハロース処理時及び再水和時に発現変動を示す遺伝子を、それぞれ 9558 及び 11997 個検出した (尤度比検定, 補正 p 値 < 0.05, BH 法)。このうち、特異的な DNA 配列への結合能を持つ転写因子のトレハロース処理時, 再水和時, 両方で発現変動した個数はそれぞれ, 5, 31 及び 76 個であり (図 3.2a), それ以外の遺伝子はそれぞれ 1249, 3662 及び 8228 個が発現変動遺伝子として検出された (図 3.2b)。トレハロース処理時及び再水和時特異的に発現変動している遺伝子数よりも両方の過程で共に発現変動している遺伝子数が非常

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.3. 結果

に多いことから、両者の過程は類似した制御機構を用いて遺伝子発現を制御していることが示された。

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.3. 結果

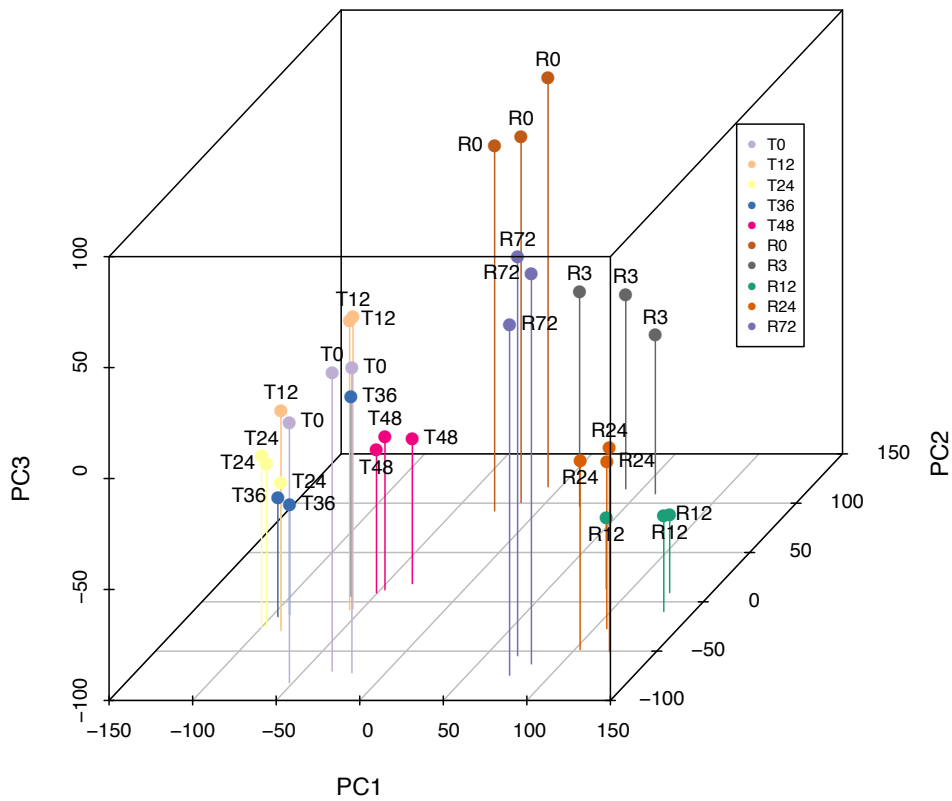


図 3.1: 全サンプルにおける主成分分析の結果

各プロットに付属する数字が処理区を表す。各軸で表現される第一、第二、第三主成分の寄与率はそれぞれ、23.6%、14.4%及び8.2%であり、全体の発現パターンの46.2%を説明する結果となる。

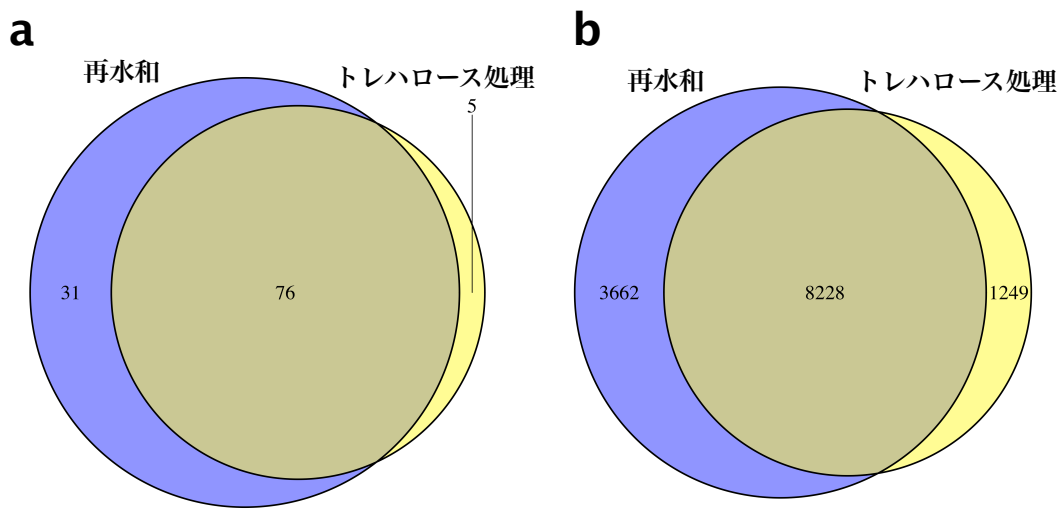


図 3.2: 転写因子及びそれ以外の遺伝子のトレハロース処理及び再水和における発現変動遺伝子の数.

a) 転写因子の発現変動遺伝子の数, b) 転写因子以外の発現変動遺伝子の数 (DESeq2, 補正 p 値 < 0.05).

3.3.2 転写制御ネットワークの推定

続いて、発現変動を示した転写因子の時系列発現量を元に、トレハロース処理及び再水和時それぞれにおいて転写制御ネットワークを推定した (図 3.3, 3.4). それぞれの転写制御ネットワークに存在した制御関係を抽出し統合転写制御ネットワークを得た (図 3.5). 先行研究において3つの転写因子間における FFL 及び FBL を持つ転写制御は遺伝子発現において重要な制御機構であることが示されている [139,140]. そこで推定した転写制御ネットワーク内のこれらの個数を算出したところ、FFL 構造が16個、FBL 構造が4個検出された. これらをさらに、FFL 構造に関しては出力に対する正負の制御の関係が一致している Coherent 型と一致しない Incoherent 型に分類したところ、それぞれ11個及び5個に分けられた (図 3.6). FBL 構造についても Feed Back 後の入力の正負によって Positive FBL 及び Negative FBL に分類したところ、それぞれ2個及び2個に分けられた (図 3.6). これら得られた構造のうち、Coherent Type FFL 構造のみランダムネットワークに対して有意に多い個数であった (Randomization test, p 値 < 0.05). Coherent Type FFL 構造は転写制御において一過的な入力をカットし、持続的な入力のみを下流に伝える役割を持つことが知られており、これはすなわちノイズカットによる正確な入力の下流への受け渡し機構として重要とされている [141]. このことから、乾燥耐性に関する転写制御において、その機構の駆動はこれら多数の Coherent Type FFL 構造により正確に入力を出力へと受け渡す機構により成り立っていると言える.

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.3. 結果

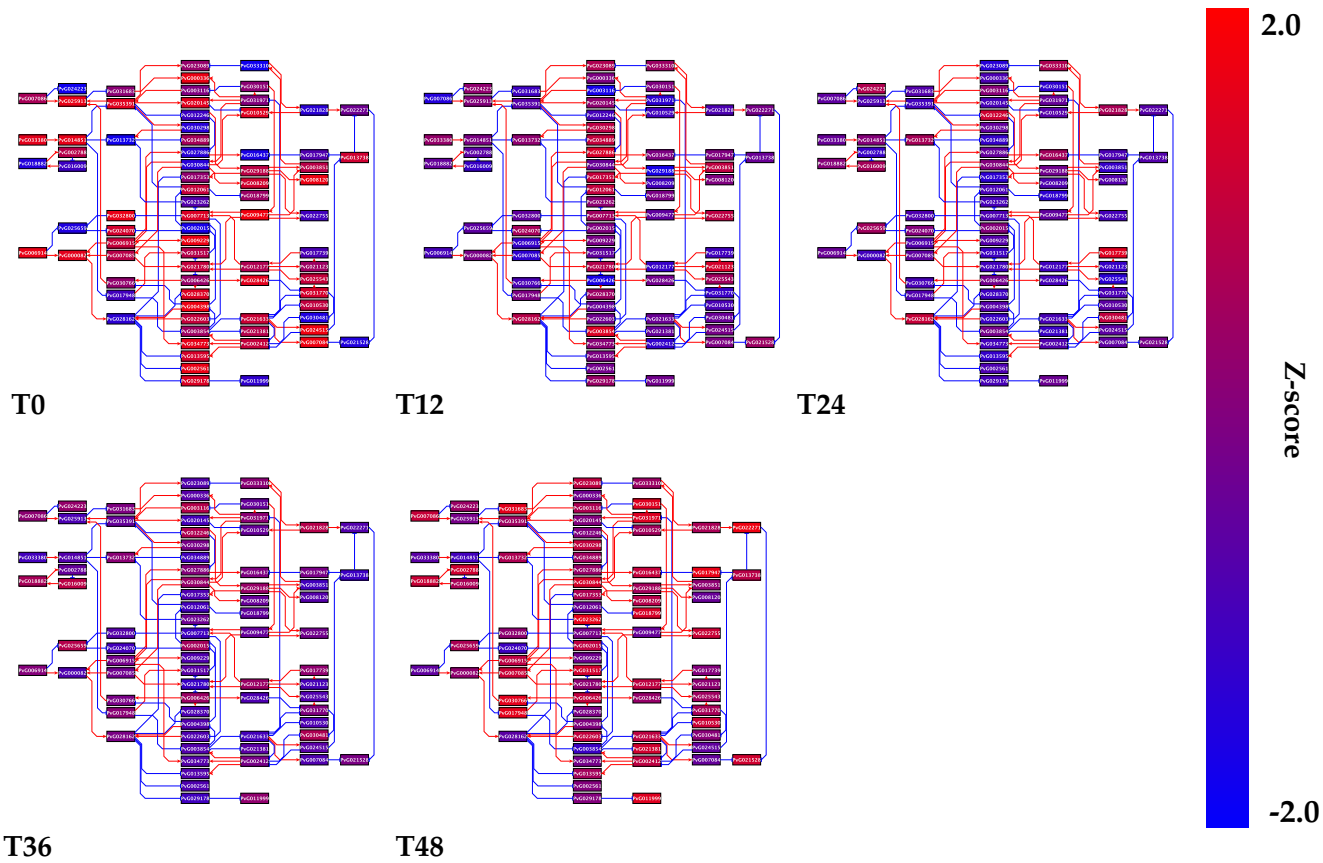


図 3.3: 推定したトレハロース処理時の転写制御ネットワーク

各四角ノードは転写因子を表し、矢印は推定された制御関係を示す。各ノードの中には転写因子の ID を示す。また、ノードは対象とする処理区における平均 RPKM の値によって配色されており、RPKM が高いほど赤色、低いほど青色で配色した。また、矢印は正、負の制御関係をそれぞれ赤、青色で示した。

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.3. 結果

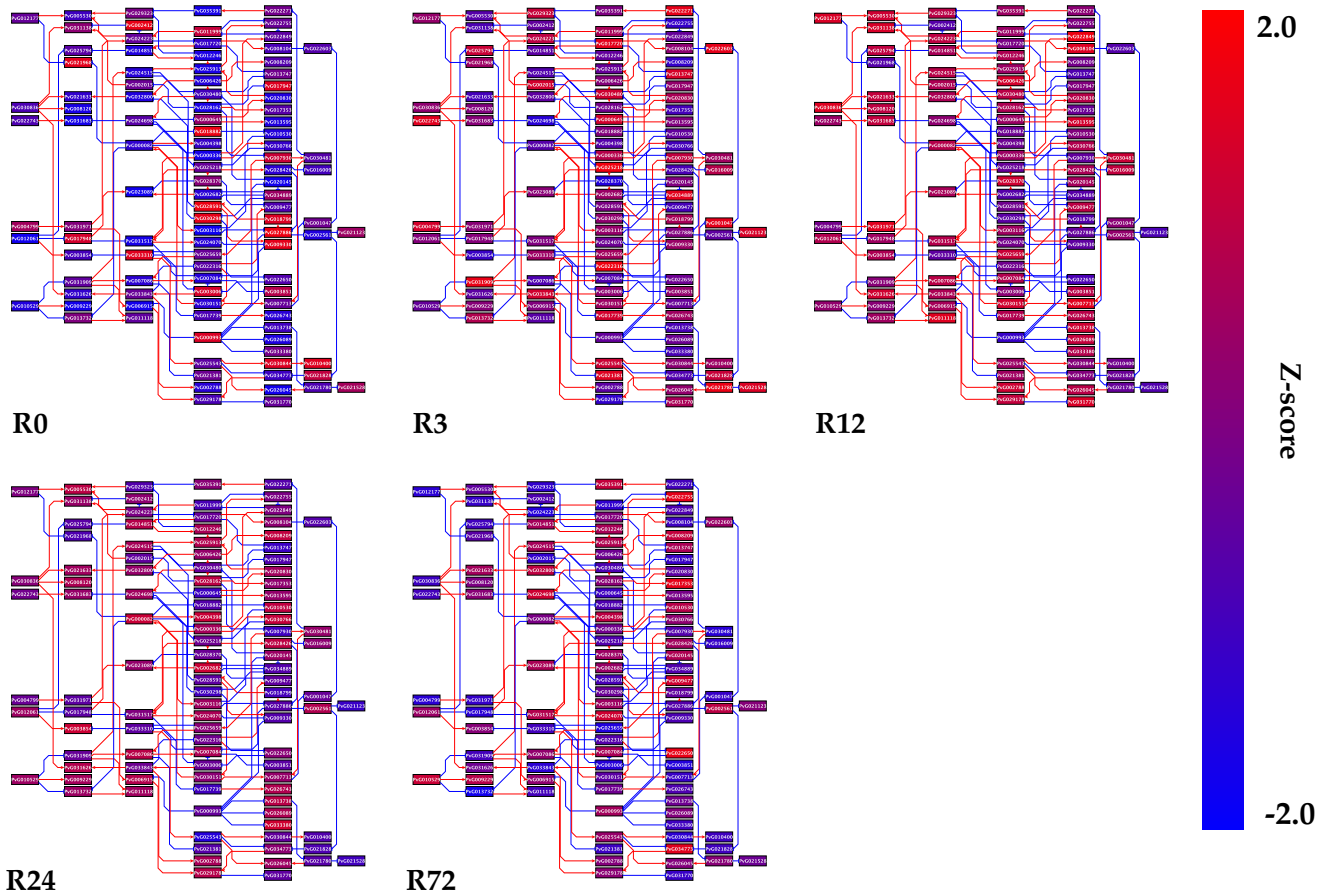


図 3.4: 推定した再水和時の転写制御ネットワーク

図内の表示方法については図 3.3 と同一とした。

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.3. 結果

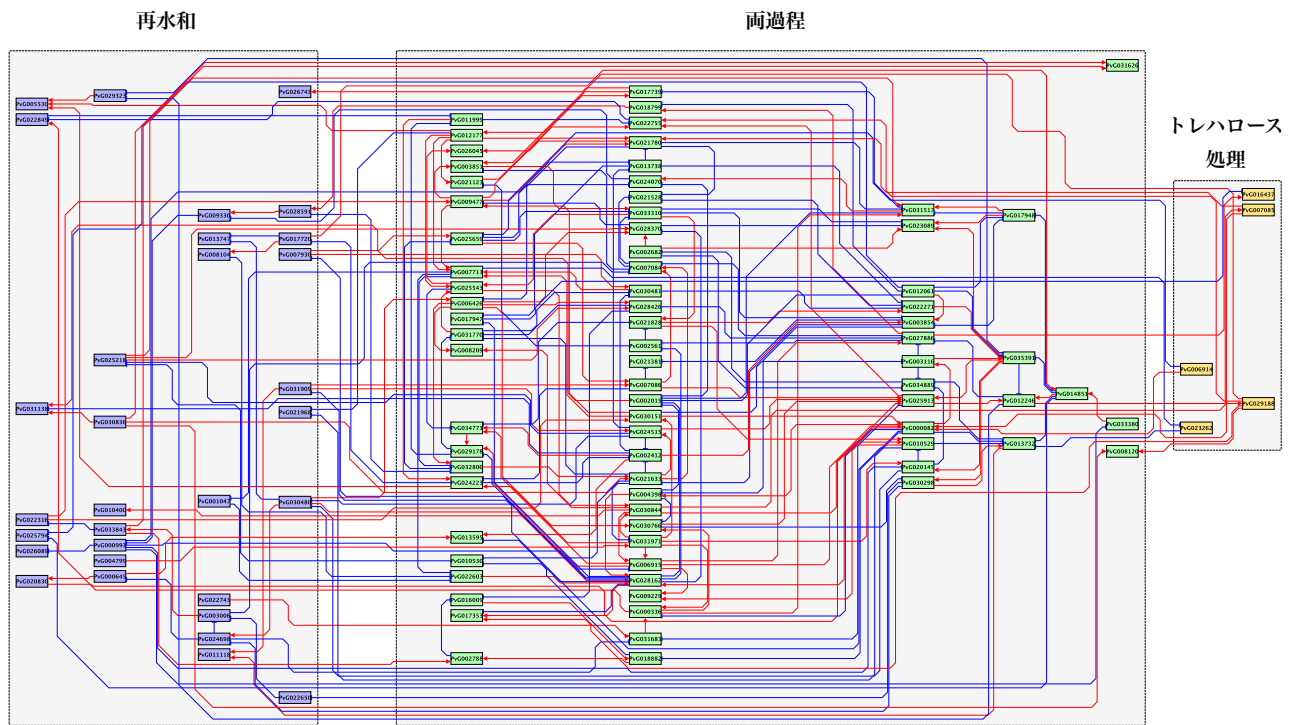


図 3.5: 推定したトレハロース処理及び再水和の統合転写制御ネットワーク

各四角ノードは転写因子を表し、矢印は推定された制御関係を示す。各ノードの中には転写因子の ID を示す。またノードはトレハロース処理、再水和及び両方で発現変動した転写因子をそれぞれ黄、青、緑色で示した。また矢印は正、負の制御関係をそれぞれ赤、青色で示した。

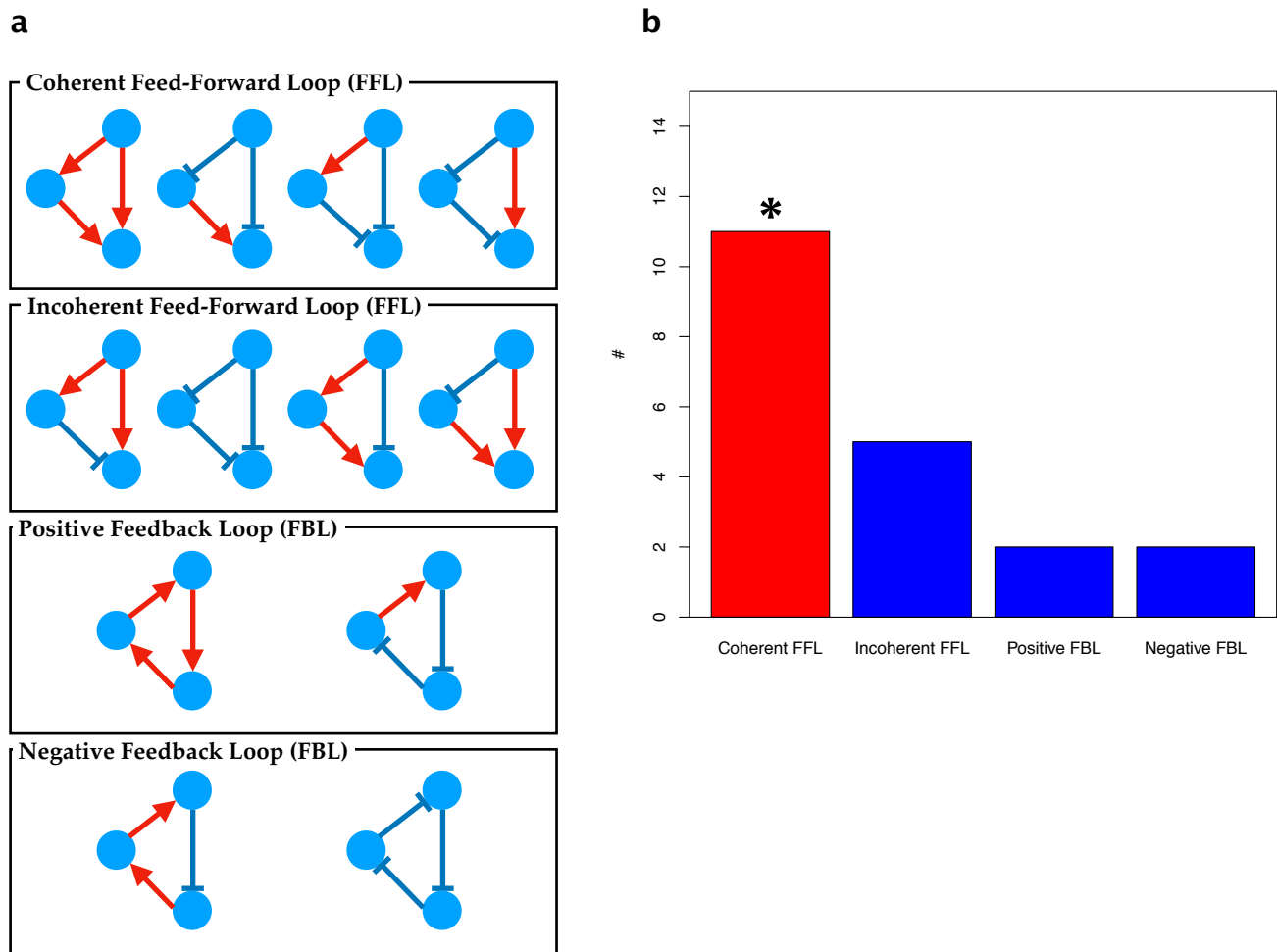


図 3.6: 遺伝子制御における重要な構造の概念図と統合転写制御ネットワークに存在する Feed Forward Loop (FFL) 及び Feedback Loop (FBL) 構造の個数

(a) FFL 及び FBL 構造の概念図. (b) 統合転写制御ネットワークに存在する FFL 及び FBL の個数. * Randomization test, p 値 < 0.05 .

3.3.3 協調発現モジュールの検出

次に、推定した転写制御ネットワーク内の各転写因子からの制御対象となる遺伝子群を同定するために、これを協調発現する遺伝子集団(モジュール)として27個のモジュールを検出した(図3.7)。これらのモジュールのうち乾燥耐性に重要なモジュールを同定するために、各モジュール内の遺伝子群に対して Gene Ontology (GO) Enrichment 解析を行ったところ、これまでに乾燥耐性への関与が報告されてきた遺伝子群の機能と合致する GO が種々のモジュールにおいて頻出していた (Fisher の正確確率検定, p 値 < 0.05)。

まず、乾燥耐性にこれまで重要とされてきたチオレドキシンは他タンパク質の酸化をシグナルとして、システイン残基でのチオール・ジスルフィド反応を介して還元を促すことで抗酸化を促す [142]。この反応に関わる Gene Ontology としては GO:0015035 (protein disulfide oxidoreductase activity) が登録されており、darkred, green, greenyellow 及び grey60 モジュールにおいて Enriched GO として検出された (p 値 = 5.04×10^{-3} , 0.0295, 1.96×10^{-3} , 5.78×10^{-4})。次に、乾燥耐性においては乾燥時に起きるタンパク質内の異性化した D 型のアスパラギン酸を元の状態に戻す PIMT の機能 [34] に関する GO として GO:0004719 (protein-L-isaspartate (D-aspartate) O-methyltransferase activity) が登録されている。この GO は darkred 及び grey60 モジュールにおいて Enriched GO として検出された (p 値 = 8.38×10^{-12} , 1.38×10^{-3})。次に、トレハロースは乾燥耐性において細胞膜の安定化に必要な物質である [118]。トレハロースに関連する GO としては、トレハロース合成に関する GO:00052992 (trehalose biosynthetic process) 及び分解に関する GO:0015927 (trehalase activity) が登録されており、それぞれ grey60 及び cyan モジュールにおいて Enriched GO として検出された (p 値 = 0.0271, 0.0163)。乾燥耐性において、再水和時には乾燥時にダメージを受けた DNA を修復することが重要であることが示されている (1.2.3.3 小々節及び 2.3.4 節参照)。DNA 修復に関する GO としては、GO:0006281 (DNA repair) が登録されており、red モジュールにて Enriched GO として検出された (p 値 = 1.68×10^{-3})。このように、これまでに報告された乾燥耐性に関与する現象と合致する GO が様々なモジュールで Enriched GO として検出されていることがわかった。

一方で、これまでに乾燥耐性への関与が示された LEA や Haemoglobin, Aquaporin に関する GO はどのモジュールにおいても検出されなかった。この原因としては、まず LEA についてはこの遺伝子の機能を表す固有の GO がいまだに登録されていないことが挙げられる。続いて、Haemoglobin と Aquaporin に関しては発現変動遺伝子として検出された個数が少なすぎるため、GO Enrichment 解析では集団内に有意に多く存在する GO を検出するという性質上、取りこぼしている可能性がある。

さらに先行研究で検証された遺伝子ファミリーである LEA, チオレドキシン, PIMT, ヘモグロビン, アクアポリン, トレハロース産生関連酵素, DNA 修復経路

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.3. 結果

(相同組替え, DNA ミスマッチ修復, ヌクレオチド除去修復及び非相同末端結合)の各モジュールへの所属数を算出した(表3.2). これらの各遺伝子ファミリーの最大所属数となるモジュールを確認したところ, それぞれのファミリーについてLEAファミリーはblackモジュール(所属数=4), チオレドキシシンファミリーはgreen及びgreenyellowモジュール(所属数=4), PIMTファミリーはdarkredモジュール(所属数=8), HBファミリーはsalmonモジュール(所属数=1), AQPファミリーはblackモジュール(所属数=2), トレハロース産生に関わる酵素はcyan, grey60, salmon, blackモジュールにそれぞれ一つずつ属していた. また, DNA修復に関する遺伝子は相同組替え及びヌクレオチド除去修復に関わる遺伝子はmagentaモジュールに所属しており(所属数=4及び4), DNAミスマッチ修復関連遺伝子はredモジュールに属していた(所属数=3). 非相同末端結合に関わる遺伝子はlightgreen, tan, brown及びdarkgreenモジュールに所属していた(所属数=1). このようにこれまでの乾燥耐性に関連することが報告されている遺伝子ファミリーに関して最大所属数を元に対応モジュールを明らかにした.

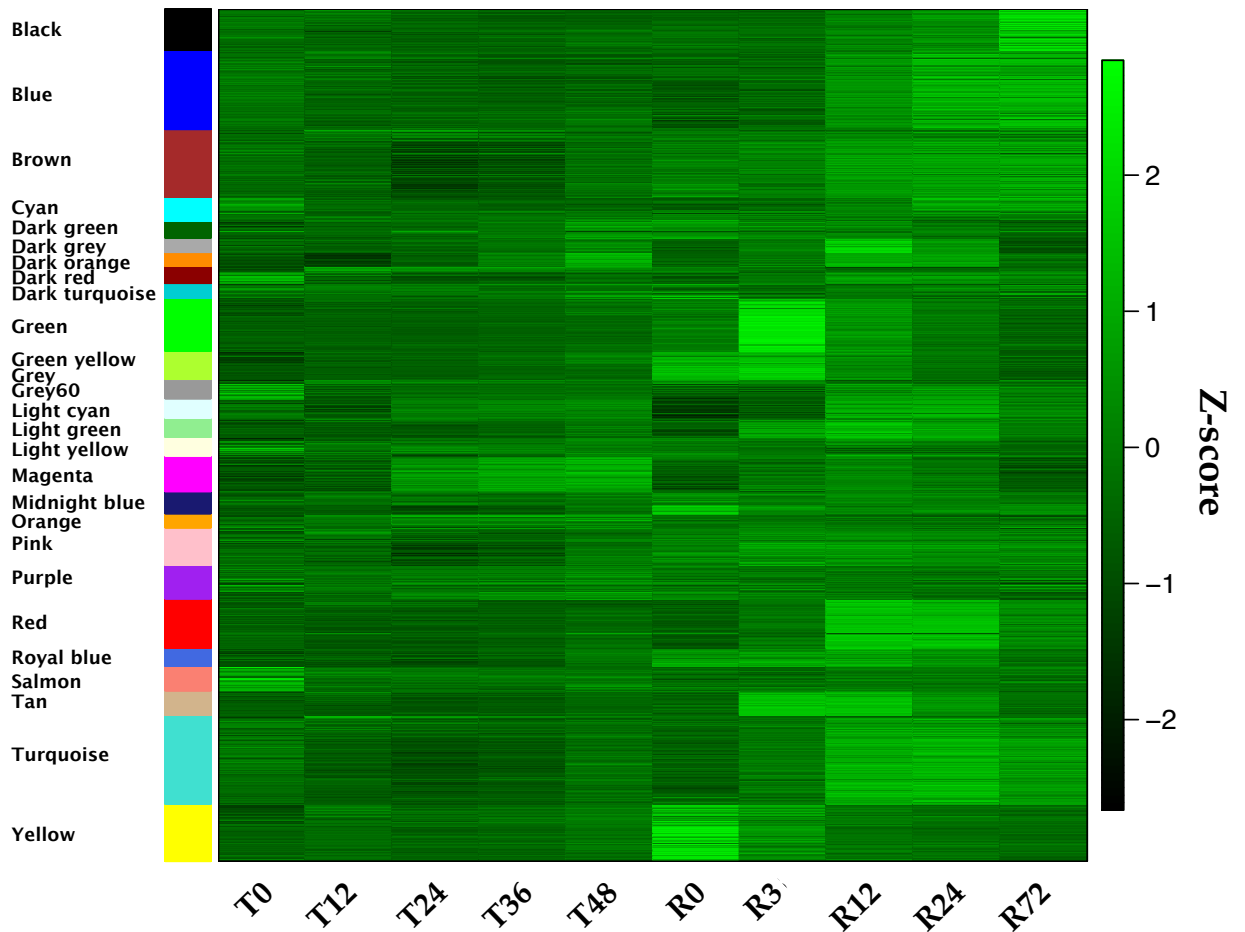


図 3.7: WGCNA により検出したモジュールの時系列発現量のヒートマップ

協調発現を示す集団をモジュールと定義，本研究ではモジュールは全て図内で表示される色名で言及する．ヒートマップの濃淡は各遺伝子ごとの時系列 RPKM から算出される Z-score の値に応じて配色した．

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.3. 結果

表 3.2: 各乾燥耐性関連遺伝子ファミリーに対する, 最大所属モジュールと所属遺伝子

遺伝子ファミリー名	最大所属モジュール	個数	配列相同な遺伝子名
LEA	black	4	PvLEA10, PvLEA13, PvLEA21, PvLEA22
TRX	green, greenyellow	4	PvTrx10, PvTrx11, PvTrx16, PvTrx17, PvTrx18, PvTrx20, PvTrx22, PvTrx23
PIMT	darkred	8	PvPimt-3, PvPimt-4, PvPimt-5, PvPimt-7, PvPimt-9, PvPimt-10, PvPimt-12, PvPimt-13
HB	salmon	1	PvHb29
AQP	black	2	Aqp1, Aqp2
Trehalose Related	cyan, grey60, salmon, black	1	Tret1, Tpp, Tps, Treh
DNA 修復 (相同組替え)	magenta	4	RAD51B, RAD51, XRCC3, DMC1
DNA 修復 (DNA ミスマッチ修復)	red	4	PMS1, PMS2, MLH1, MLH3
DNA 修復 (ヌクレオチド除去修復)	magenta	3	MNAT1, CDK7, GTF2H4
DNA 修復 (非相同末端結合)	lightgreen, tan, brown, darkgreen	1	XRCC6, XRCC5, LIG4, PRKDC

3.3.4 転写制御ネットワークと協調発現モジュールの統合

続いて、転写因子のモジュール内遺伝子群の上流領域への結合予測として Motif Enrichment 解析 (CLOVER [86], Randomization test, p 値 < 0.05) 及び転写因子の発現変化がモジュール内の遺伝子群の発現変化に対して Granger 因果を持つかを解析し (Granger 因果性検定, 補正 p 値 < 0.05 , BH 法), 両方の解析で支持された転写因子とモジュール間の関係を制御関係として検出した (図 3.8, 表 3.3). この結果より, 全てのモジュールは HSF(PvG018882) を上流に持つことが明らかとなり, 先行研究同様に HSF が乾燥耐性機構を制御する上で重要な転写因子であった.

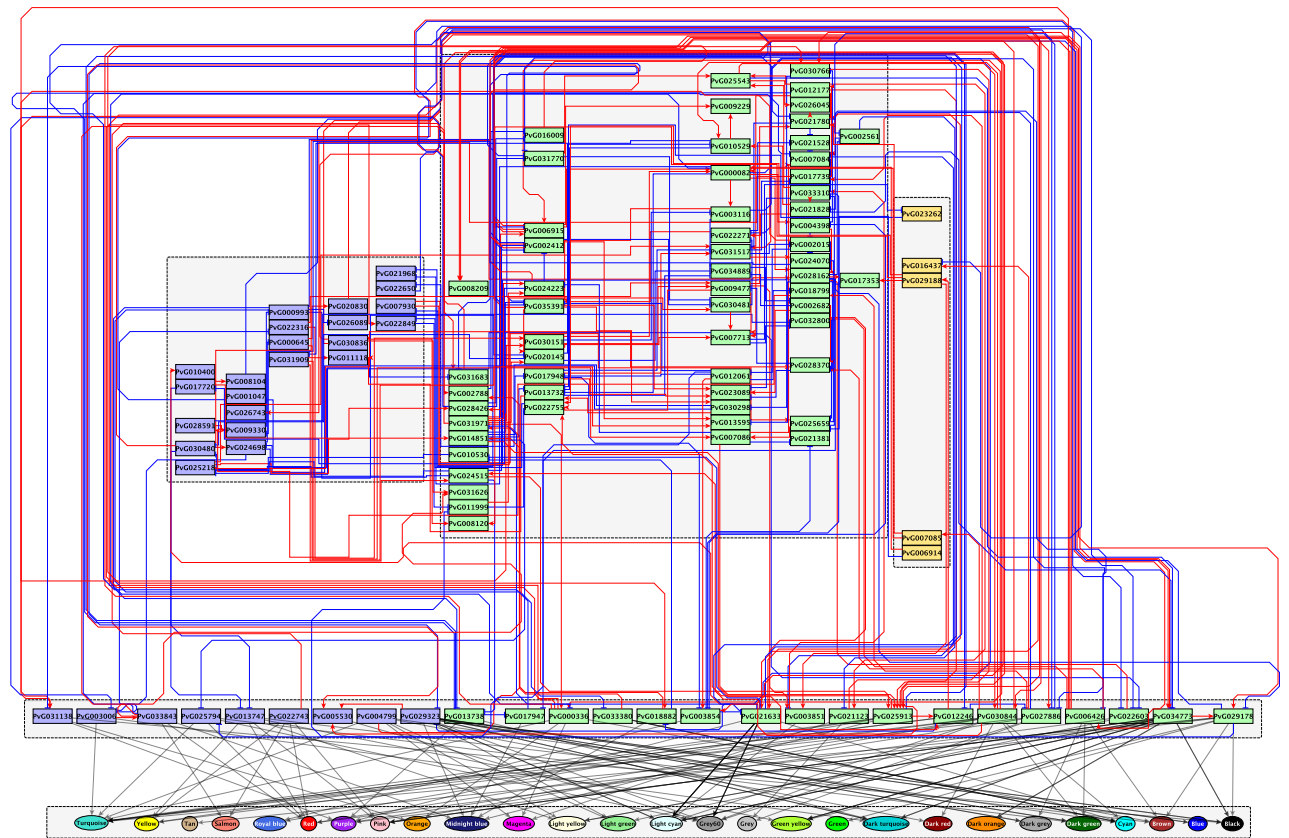


図 3.8: 転写因子と協調発現モジュールの制御関係をまとめた遺伝子制御ネットワーク

四角ノードは転写因子、丸ノードはモジュールを表す。矢印は推定された制御関係を表す。転写因子のノード及び転写因子間の制御関係の配色は図 3.5 と同一である。また、転写因子からモジュールへの制御関係は灰色で示した (CLOVER, p 値 < 0.05 かつ Granger 因果検定の補正 p 値 < 0.05 , BH 法)。

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.3. 結果

表 3.3: 各モジュールを直接制御する転写因子及び HSF(PvG01882) による制御の有無

HSF 依存性有りとは対象となるモジュールの上流を遡った時に HSF が存在する場合を指す。

モジュール	モジュールを直接制御をする転写因子	HSF (PvG01882) 依存性
black	PvG029178; PvG034773; PvG029323; PvG031138	TRUE
blue	PvG029323	TRUE
brown	PvG029178; PvG006426; PvG000336; PvG013747	TRUE
cyan	PvG030844; PvG021633; PvG033380; PvG004799;	TRUE
darkgreen	PvG006426; PvG030844; PvG029323; PvG025794; PvG005038	TRUE
darkgrey	PvG006426; PvG029323	TRUE
darkorange	PvG034773; PvG003458; PvG005038	TRUE
darkred	PvG030844; PvG004799; PvG005038	TRUE
darkturquoise	PvG013738; PvG034773; PvG003006; PvG029323	TRUE
green	PvG029178	TRUE
greenyellow	PvG029178; PvG034773; PvG013747; PvG005038	TRUE
grey	PvG027886; PvG021123;	TRUE
grey60	PvG027886; PvG030844; PvG021633; PvG029323; PvG025794; PvG005530	TRUE
lightcyan	PvG030844; PvG021633; PvG003851; PvG034773; PvG029323	TRUE
lightgreen	PvG022603; PvG029323; PvG004799	TRUE
lightyellow	PvG034773; PvG004799; PvG013747; PvG003458	TRUE
magenta	PvG021123; PvG000336; PvG004799	TRUE
midnightblue	PvG025913; PvG018882; PvG017947; PvG033843; PvG029323	TRUE
orange	PvG000336; PvG031138;	TRUE
pink	PvG030844; PvG034773; PvG022743; PvG013747	TRUE
purple	PvG027886; PvG029178; PvG021633; PvG013738; PvG031138	TRUE
red	PvG022603; PvG030844; PvG033843; PvG022743; PvG013747; PvG025794; PvG005038	TRUE
royalblue	PvG025913; PvG003854;	TRUE
salmon	PvG033843; PvG022743; PvG005038	TRUE
tan	PvG029178	TRUE
turquoise	PvG025913; PvG027886; PvG029178; PvG030844; PvG012246; PvG000336; PvG003006; PvG004799; PvG025794; PvG003458; PvG005038	TRUE
yellow	PvG025913; PvG030844; PvG022743; PvG013747; PvG005038	TRUE

3.3.5 ショウジョウバエ遺伝子制御ネットワークとの比較

ここから、乾燥耐性を持つ Pv11 細胞特有の転写制御構造を探索するために、modEncode プロジェクトにて公開されているショウジョウバエの転写制御ネットワーク [61] との比較解析を行った (図 3.9). ショウジョウバエは乾燥耐性を持たないため、乾燥耐性制御機構の理解に向けて Pv11 細胞の乾燥耐性関連転写制御ネットワーク内からショウジョウバエとの間で配列相同性が認められなかった転写因子を中心に縮約を行った (3.2.7 小節参照, 図 3.10). これにより、乾燥耐性を持つ Pv11 細胞のみが持つ縮約された転写制御ネットワーク内における最上流となる転写因子は PvG003116 であることが明らかになった. また、この PvG003116 は nuclear transcription factor Y subunit gamma-like (NF-YC, Accession Number : XP_029711307.1) との間で配列相同性が確認された (blastp, E-value = 1.0e-75). さらに、得られた縮約後の転写制御ネットワークには3つの FFL 構造と3つの FFL 構造が互いに転写因子を共有しあった構造 (Triple FFL) が存在し、これらを一つのブロックとしてまとめた (図 3.11). これにより、NF-YC の下流に Triple FFL と FFL3 の間、FFL2 と FFL3 の間、Triple FFL と FFL2 と FFL3 の間、そして HSF と FFL2 と Triple FFL の間に4つの Positive FBL 構造が存在することが明らかになった. 加えてこれらの制御関係の下流で LEA, チオレドキシン, アクアポリン, ヘモグロビン, トレハロース産生及び DNA 修復モジュールが制御を受けていた.

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.3. 結果

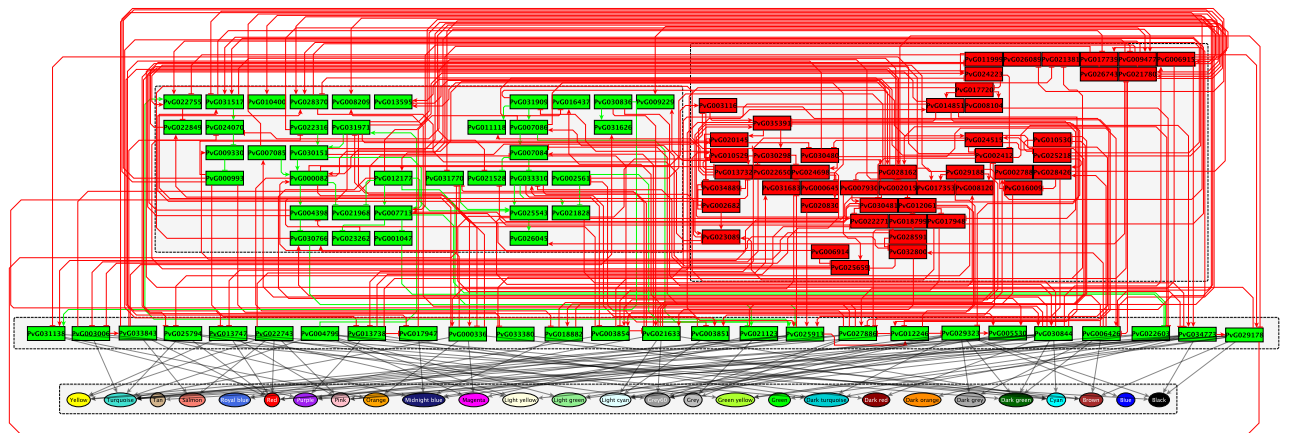


図 3.9: ショウジョウバエと Pv11 細胞の転写制御ネットワークの比較解析結果

四角ノード及び丸ノードは転写因子及びモジュールを表す。矢印は制御関係を表す。四角ノードはショウジョウバエとの間で配列相同性の確認された転写因子を緑色 (blastp, $E\text{-value} < 1.0e-15$), 確認されなかった転写因子を赤色で示している。矢印の配色は, ショウジョウバエ及び Pv11 細胞の転写制御ネットワークのいずれにも確認された制御関係を緑色, Pv11 細胞のみで確認された制御関係を赤色で表示した。

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.3. 結果

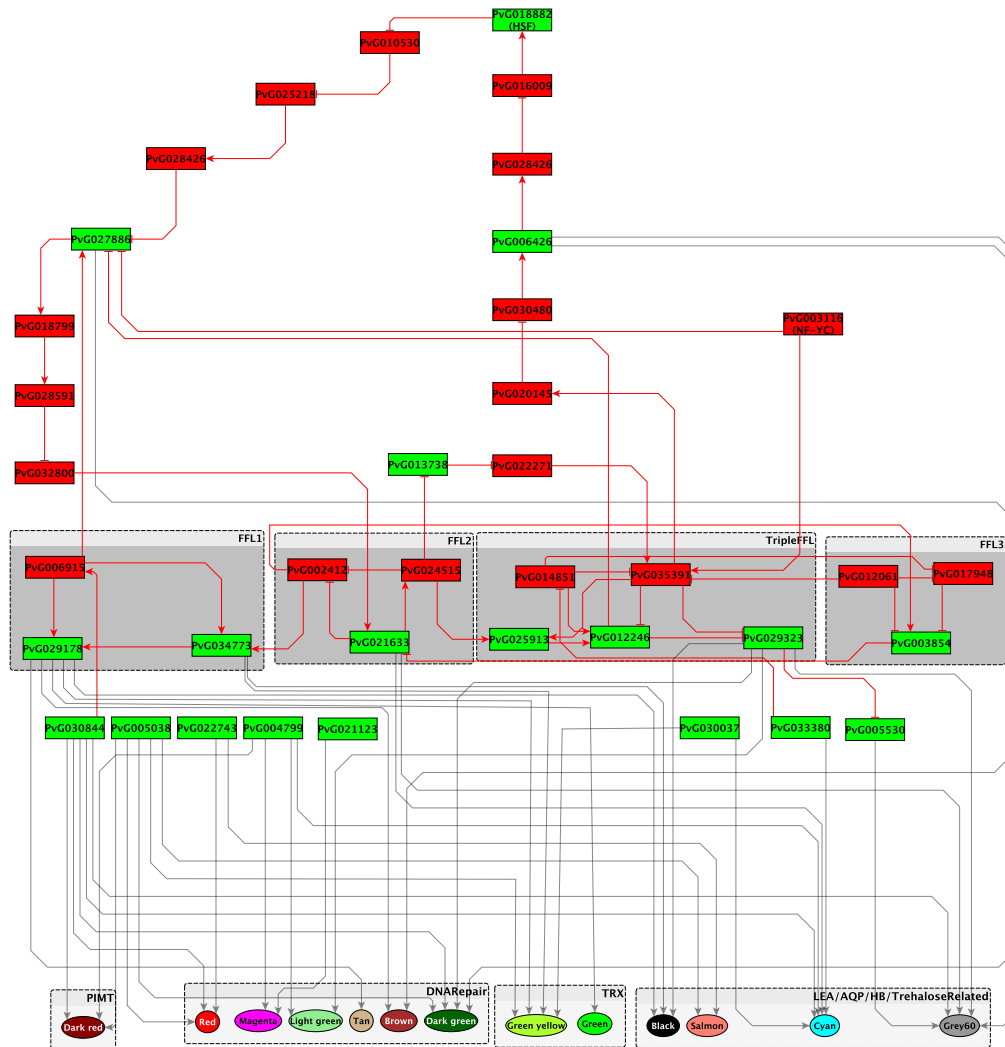


図 3.10: ショウジョウバエと Pv11 細胞の転写制御ネットワークの比較解析結果から得られた縮約ネットワーク

四角ノード及び丸ノードは転写因子及びモジュールを表す。矢印は制御関係を表す。四角ノードはショウジョウバエとの間で配列相同性の確認された転写因子を緑色 (blastp, E-value < 1.0e-15), 確認されなかった転写因子を赤色で示している。また, FFL を構成する転写因子群を FFL1, FFL2, FFL3, TripleFFL ブロックとしてまとめている。矢印の配色は, ショウジョウバエ及び Pv11 細胞の転写制御ネットワークのいずれにも確認された制御関係を緑色, Pv11 細胞のみで確認された制御関係を赤色で表示した。モジュールについては各ファミリー遺伝子の所属数をもっとも多かったモジュールをファミリーごとにそれぞれまとめて表示している。

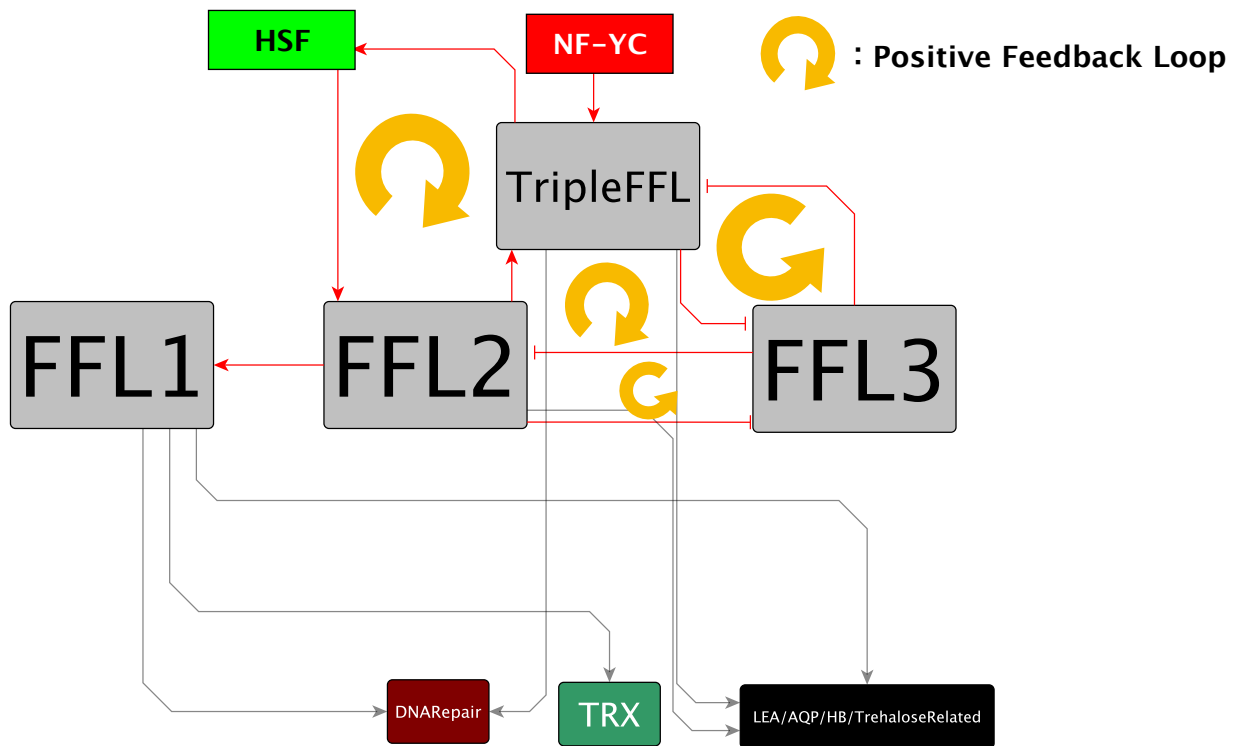


図 3.11: 最終縮約遺伝子制御ネットワーク

このネットワーク内には Triple FFL と FFL3 の間, FFL2 と FFL3 の間, Triple FFL と FFL2 と FFL3 の間, そして HSF と FFL2 と Triple FFL の間に 4 つの Positive FBL 構造が存在する.

3.3.6 NF-YCから乾燥耐性関連遺伝子群への制御関係に関する妥当性検証

今回遺伝子制御ネットワーク推定に基づきNF-YCが乾燥耐性関連遺伝子群を制御する可能性があるが、統計的な推定においては偽陽性が存在するため、今回の示唆が偽陽性の可能性も考えられる。そこで、この示唆の妥当性を検証するためにNF-YCを最上流とした乾燥耐性関連遺伝子群を直接制御する転写因子(PvG029178)への制御関係が少なくとも一つ存在する確率を一本あたりの制御関係の真の確率 p に対する関数として(3.5)式を用いて求めた(図3.12)。また、本研究で用いたBTNETのベンチマーク問題に対するPrecisionの最高値が0.359であったことから、この値を前述の p に代入したところ、0.997と算出された。PvG029178から乾燥耐性関連遺伝子群への制御関係はFDR < 0.05で検出している。これらのことを統合するとBTNETの最高精度での解析が行えていたとするならば、本研究で示唆した仮説であるNF-YCが乾燥耐性関連遺伝子群を制御するという仮説は、 $0.997 \times 0.95 = 0.947 = 94.7\%$ 確からしいと言える。

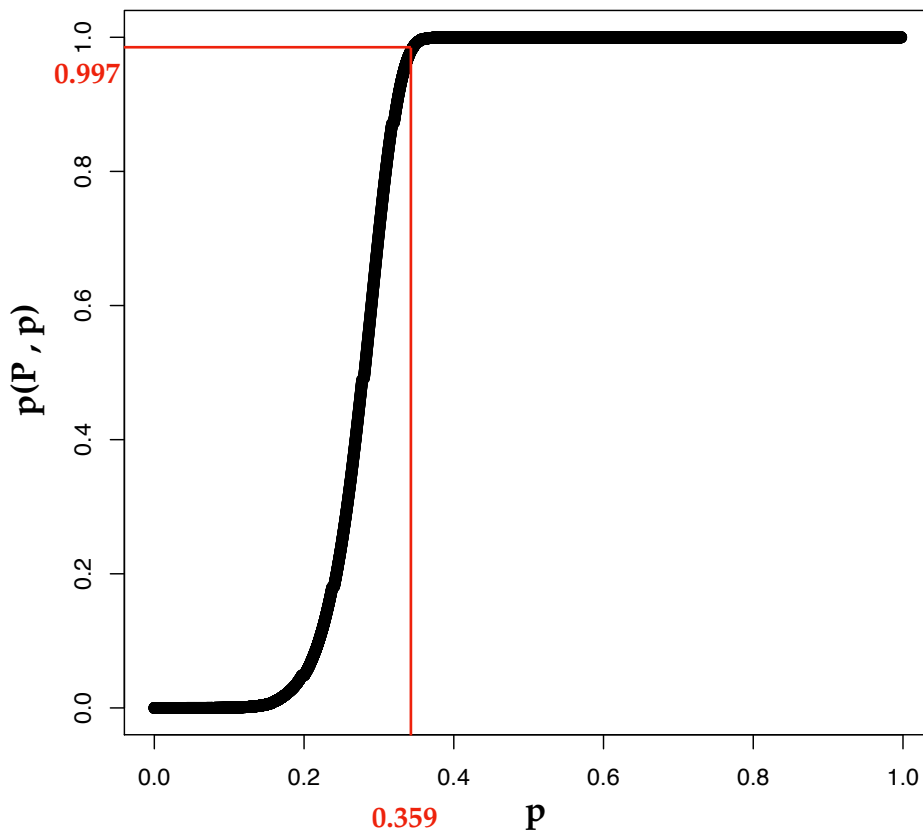


図 3.12: NF-YC を最上流とした乾燥耐性関連遺伝子群を直接制御する転写因子 (PvG029178) への制御関係が少なくとも一つ存在する確率

始点 v_s を NF-YC, 終点 v_f を乾燥耐性関連遺伝子群を直接制御する転写因子 (PvG029178) とした場合の (3.5) 式における一本あたりの制御関係が真である確率 p に対する v_s から v_f へのパスが少なくとも一つ存在する確率 $p(\mathbf{P}, p)$ の関係を黒線で示す. BTNET でのベンチマーク問題に対する Precision と Recall の調和平均を最大化する Precision が 0.359 であり, この値を p とした場合に $p(\mathbf{P}, p)$ は 0.997 と算出された (赤線).

3.4 考察

3.4.1 再水和直後におけるミトコンドリア転写活性と乾燥耐性機構の関係

今回新規に採取した RNA-seq データにおいて、R3, R12, R24 において他のサンプルに比べて顕著にマッピング率が低かった (表 3.1)。これはマッピング時に対象としているネムリユスリカ染色体由来の転写物ではない由来不明の転写物が多数存在することになる。ここで、Pv11 細胞内において RNA の転写が生じる対象は染色体以外にミトコンドリアゲノムが存在することに注目し、今回採取した RNA-seq データを染色体ゲノムとミトコンドリアゲノム [143] の両方にマッピングし、マッピング率を算出した (図 3.13)。この結果、マッピング率の低かった R3, R12, R24 の処理区においてミトコンドリアゲノムにマッピングされるリードの割合が他の処理区に比べて顕著に高かった。このように再水和直後の処理区にのみミトコンドリア由来 RNA の割合が顕著に高いという事実は、ミトコンドリアがエネルギー産生のものであるということを考慮すると、再水和という処理によって Pv11 細胞は能動的に生き返るためのエネルギー産生をするという仕組みが存在する可能性がある。乾燥耐性において、このような再水和直後における能動的なミトコンドリアゲノムの転写活性が報告された事例はないため、この仕組みの存在については今後さらに研究を進めることで生き返るための能動的な仕組みの発見できる可能性がある。

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.4. 考察

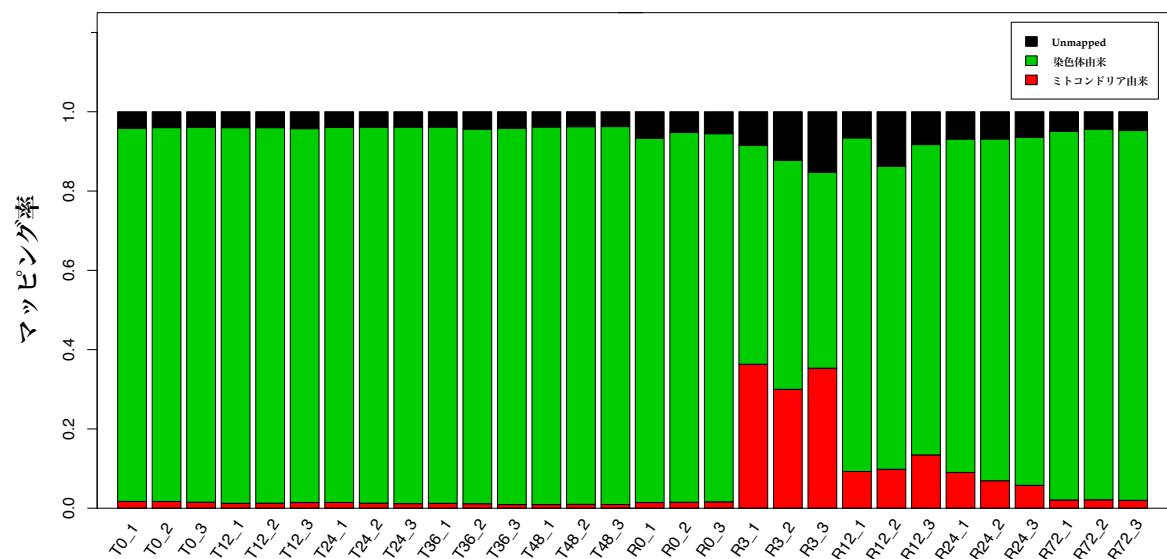


図 3.13: 各サンプルの染色体及びミトコンドリアゲノムへのマッピング率

赤色がミトコンドリアにマッピングされたリード，緑色が染色体にマッピングされたリード，黒色がマッピングされなかったリードの割合を示す．R3においてミトコンドリアへのマッピング率が極端に上昇していた．

3.4.2 GO解析結果を利用したエフェクター遺伝子による新たな乾燥耐性機構発見の可能性

本研究ではGO解析の結果は最終的に縮約した遺伝子制御ネットワークを構築するための事前情報としては利用せずに、先行研究で得られている乾燥耐性に関連する機能情報が協調発現を示す集団であるモジュールに特徴的に存在するかどうかの検証のみに用いた。この結果、これまでにネムリユスリカの乾燥耐性において報告されてきた機能情報の多くに関連するGOがいずれかのモジュールにおいてEnriched GOとして得られていることを確認した。しかし、今回注目したGO以外にも、より多くの様々なGOが各モジュールのEnriched GOとして検出された。これらのEnriched GOの中にはいまだに乾燥耐性との関連が報告されていないものも多く存在し、モジュール構成遺伝子の時系列発現のパターンと統合してより詳細に調査することにより、乾燥耐性における新たな分子生物学的なメカニズムの発見に結びつく可能性がある

3.4.3 NF-YCによる乾燥耐性制御の可能性

今回の解析により乾燥耐性機構を最上流で制御していると推定されたNF-YCはnuclear factor Y (NF-Y) 転写因子を構成するサブユニットである [144]。NF-Y自体はNF-YA, NF-YB, NF-YCの3つのサブユニットから構成される。NF-YCは、これらNF-YA, NF-YBとの相互作用を介してゲノム上のCCAATモチーフへ結合し、種々の遺伝子の発現制御を行うことが報告されている。先行研究では、NF-YBを恒常発現するシロイヌナズナの株を作出することで、野生型ではしおれてしまう8日間の干ばつストレス後にもしおれることなく通常の状態を維持する [145]。またNF-YCを過剰発現させたイネは、野生型では完全にしおれてしまう10日間の完全な水の遮断処理においてもしおれずに脹れた状態 (turgid) を保たれる [146]。このように、NF-Yは植物において広く干ばつストレス耐性獲得に必須な転写因子として同定されている一方で、このような報告はネムリユスリカを始めとした動物においては報告はない。今回の結果は動物においてもNF-Yが乾燥耐性に関わることを示唆する。またネムリユスリカの虫体においてはNF-YCの発現は乾燥及び再水和に応じて発現上昇を示す結果が得られている (Pv.00526の乾燥0時間に対する乾燥24時間後のfold change = 1.22, 乾燥48時間後に対する再水和3時間後のfold change = 7.22, [19])。さらに興味深いことにネムリユスリカの近縁種で乾燥耐性を持たないヤモンユスリカはNF-YCと配列相同性を持つ遺伝子としてPn.05756を持ち (blastn, E-value = 9e-32), 乾燥後の発現上昇はネムリユスリカの虫体とそれほど変わらない (乾燥0時間に対する乾燥24時間後のfold change = 1.21, [19])。またショウジョウバエにおいてもネムリユスリカにけるNF-YCと配列相同な転写因子である nuclear factor Y-box C 195 (FBtr0070974) が確認されているが (tblastn, E-value = 1.35e-38), 乾燥時における発現変動は知られていな

3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定3.4. 考察

い。このようにNF-YCはショウジョウバエ、ネムリユスリカ及びヤモンユスリカという昆虫において広く保存されているが、今回の結果を鑑みるとPv11細胞、そしてその由来の個体であるネムリユスリカのみ乾燥耐性の獲得に貢献すると考えられる。

3.4.4 乾燥耐性を持たないヤモンユスリカが乾燥耐性を持たない理由

それでは、ヤモンユスリカはNF-YCをネムリユスリカと同等に乾燥に応じて発現上昇させられるのに、なぜ乾燥耐性を持っていないのだろうか？非常に興味深いことに、NF-YCから乾燥耐性関連遺伝子を含むモジュールに至るまでに必要なPvG021633(FFL2からLEA/AQP/HB/TrehaloseRelatedモジュールへの接続部分)、PvG024515(FFL2からFFL1への接続部分)、PvG029323(TripleFFLからDNARepair及びLEA/AQP/HB/TrehaloseRelatedモジュールへの接続部分)と配列相同性を持つ遺伝子がヤモンユスリカに存在しないことが明らかとなった (blastn, E-value < 1.0e-5, [19])。このことから、ヤモンユスリカが乾燥耐性を持っていない理由は、LEAやチオレドキシシンなどの乾燥耐性を発揮するための遺伝子への制御に必要な転写因子を欠損しているためだと考えられる。

3.4.5 Positive Feedback Loop構造を利用した乾燥耐性制御機構

今回、Pv11細胞の転写制御ネットワークを推定することで、NF-YC下流に多数のPositive FBLが存在することが明らかになった (図3.11)。Positive FBLは転写制御において、入力に応じた出力のON/OFFを制御する構造であることが報告されている [139]。このことから、乾燥に応じたNF-YCからの入力に対する乾燥耐性関連遺伝子の発現は、これら多数のPositive FBLによってON/OFFの制御を受けることで乾燥耐性制御機構が構築されていると考えられる。

これらの結果を総括して考えると、Pv11細胞においてはトレハロース処理や乾燥に応じて乾燥耐性関連遺伝子の発現がONになり乾燥しても死なない状態に、再水和によって元の状態に戻るといのように、Pv11細胞はこれらのPositive Feedback Loopを利用することで無代謝状態と通常状態を切り替える制御を行うことで乾燥耐性自体が成り立っている可能性がある (図3.14)。

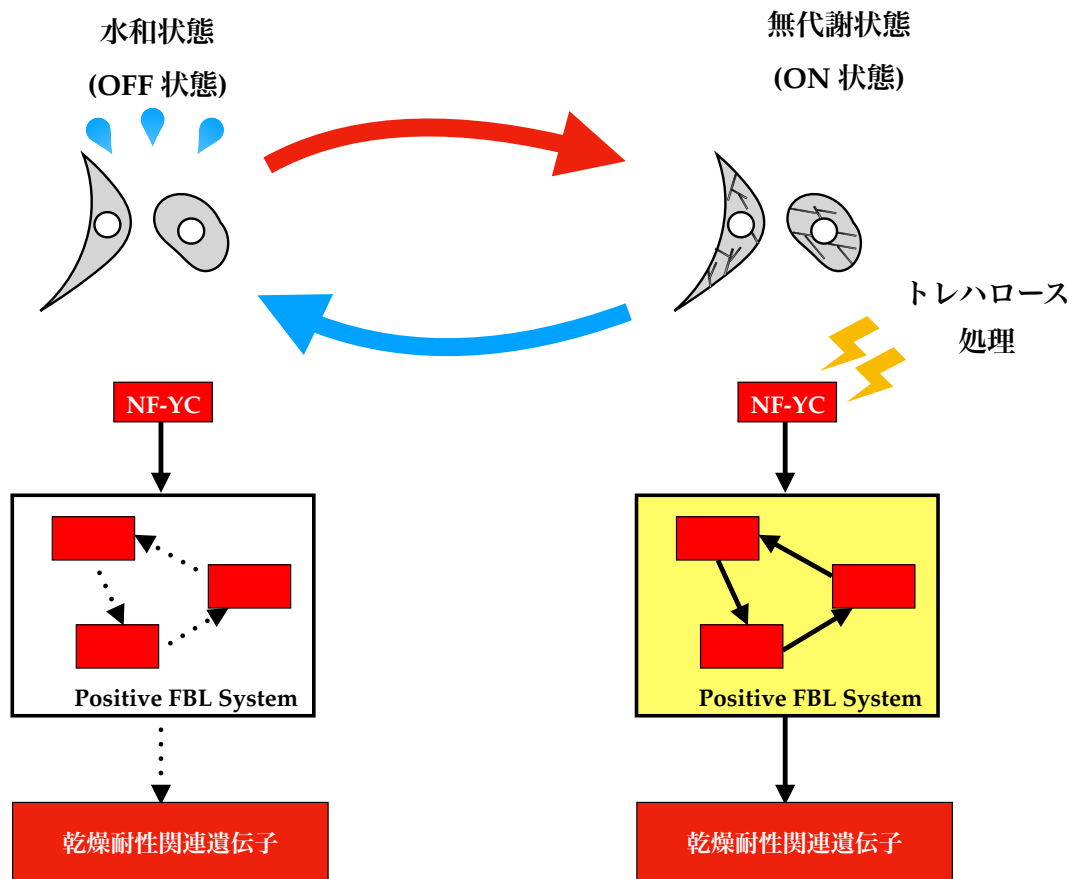


図 3.14: 本研究で示唆した Pv11 細胞における乾燥耐性 ON/OFF 制御機構の概念図

本研究で推定した Pv11 細胞の乾燥耐性に関する遺伝子制御ネットワークにおいては 4 つの Positive FBL が存在した (図 3.11). Positive FBL は遺伝子制御において入力の強弱に応じた出力の ON/OFF を制御する構造であることが報告されている [139]. この知見と本研究の結果を統合して考えると, トレハロース処理や乾燥に応じて乾燥耐性関連遺伝子の発現が ON になり Pv11 細胞は乾燥しても死なない状態に, 再水和によって元の状態に戻るといのように, Pv11 細胞はこれらの Positive Feedback Loop を利用することで無代謝状態と通常状態を切り替える制御を行うことで乾燥耐性自体が成り立っている可能性がある.

第4章 結論

4.1 結言

本研究では、Pv11細胞の乾燥耐性に関与する遺伝子群およびそれらを制御する転写制御ネットワークの推定に基づく乾燥耐性制御機構に関する知見を得ることを目的とした。そして、Pv11細胞の乾燥耐性獲得期においては生体障害緩和に関わる遺伝子群、再水和期においてはDNA修復に関わる遺伝子群が関与する可能性を示した。これら遺伝子群の上流の制御関係を推定することでPv11細胞の乾燥耐性制御機構を最上流で制御する転写因子としてNF-YCが存在すること、そしてNF-YCから乾燥耐性関連遺伝子群の制御には複数のFeed Forward Loop制御関係とそれらの間のPositive Feedback Loop制御に基づく安定したON/OFF制御機構により乾燥耐性制御機構が構築されている可能性を示した。

4.2 Pv11細胞の乾燥耐性獲得および再水和復帰を支える遺伝子群の同定

2章では、Pv11細胞の乾燥耐性獲得および再水和復帰の機構を明らかにすることを目的として、各過程において発現変動遺伝子の検出およびその機能解析を行なった。その結果、乾燥耐性獲得期においては個体レベルの乾燥耐性においても重要性が示されているLEA、チオレドキシン、PIMT遺伝子といった生体障害の緩和に関わる遺伝子の一部の高発現に加え、種々の転写物や脂質膜への障害を修復するストレス応答遺伝子(NMD Protein, glutathione S-transferase, ATP-dependent RNA helicase および DEAD box ATP-dependent RNA helicase) や生体障害となる化学物質の無毒化を行なう遺伝子(UDP-glucuronosyltransferase や carboxylesterase) がPv11細胞のトレハロース処理時特異的に顕著に発現上昇することが確認された(2.3.3および2.3.5小節参照)。

再水和期においては、相同組替え、ヌクレオチド除去修復および非相同末端結合に基づくDNA修復機構に関与する遺伝子群の顕著な発現上昇が確認され、この中でも特に非相同末端結合に関する遺伝子は個体での顕著な発現上昇が見られるPv11細胞特異的に発現上昇することが明らかとなった(2.3.4および2.3.5小節参照)。

4 結論4.3. Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定

これらの結果を統合して考えることで、Pv11細胞における細胞レベルの乾燥耐性は個体レベルの乾燥耐性よりも乾燥耐性獲得期においてはより広い範囲の障害に対する対策を行なうための遺伝子の準備を行なう機構を備えていること、再水和期においては個体におけるDNA修復機構に加えて早急にDNAへの障害を修復する非相同末端結合機構を備えていることを示唆した。

本研究で得られたこれらの知見は、これまでに未知だったPv11細胞の細胞レベルの乾燥耐性機構に関する初めての分子生物学的な知見であり、本機構を構成する遺伝子群の候補を示した。

4.3 Pv11細胞の乾燥耐性を支えるマスター転写因子と転写制御機構の推定

3章では2章で得られた知見も含めて、これまでに乾燥耐性に重要であることが示された遺伝子群の制御機構の解明を目指し、トレハロース処理および再水和時における時系列遺伝子発現量を元に遺伝子制御ネットワークの推定を行なった。推定した遺伝子制御ネットワークを乾燥耐性を持たないショウジョウバエの公知の転写制御ネットワークと比較することにより、植物において広く干ばつストレス耐性に重要であることが報告されているNF-YCが乾燥耐性関連遺伝子群の最上流転写因子として働いている可能性を示した(3.3.5小節参照)。さらに、これら転写因子と乾燥耐性関連遺伝子群の間の制御関係に注目することで、ノイズ除去に重要なFeed Forward Loop制御構造と、その間における複数のPositive Feedback Loop制御が存在することを明らかにした。Positive Feedback Loop制御構造は入力強度に応じた出力のON/OFFの安定した切り替えに重要であるということが報告されている。この先行研究と本結果を統合して考えることで、トレハロース処理や乾燥に応じて乾燥耐性関連遺伝子の発現がONになりPv11細胞は乾燥しても死なない状態に、再水和によって元の状態に戻るというように、細胞レベルの乾燥耐性はPositive Feedback Loopを利用することで無代謝状態と通常状態を切り替える制御により成り立つ可能性を示唆した(図3.11)。

本研究により得られた知見は、乾燥耐性においていまだ明らかにされていなかった遺伝子制御の機構に関する知見であり、今回NF-YCを初めとして乾燥耐性に重要な役割を果たす可能性がある転写因子とその制御関係に関する候補を示すものである。これらの知見は、今後の乾燥耐性制御による別生物種への乾燥耐性付与による保存が難しい生鮮食品などの長期保存を可能にする技術の開発などを目指した研究において、今回同定したNF-YCやPositive Feedback Loop制御機構を対象とする生物へ遺伝子導入により組み込むことを提案するという点で重要な役割を果たすと考えられる。

4.4 今後の展望

まず初めに、本解析によって乾燥耐性を構築するために十分な転写制御の関係がシステムレベルで示唆されたわけだが、これらのうち乾燥耐性機構を成り立たせるために実際にはどの転写制御が必要なのかどうかを明らかにする必要がある。近年、Pv11細胞に対してはRNAiを用いた knock down による実験系が構築された [147]。この系を利用することで、今回提案した NF-YC や Positive FBL を構成する転写因子を Knock Down した際に Pv11 細胞の乾燥耐性が失われるのかを検証することで必須な転写因子とその転写制御構造を明らかにできると考えられる。

次の展望として、ネムリユスリカの虫体と Pv11 細胞の乾燥耐性の違いを明らかにすることが考えられる。現在確認されていることとして、Pv11 細胞は再水和後の蘇生率が 20% 以下にとどまっている一方で、ネムリユスリカ虫体の再水和後の蘇生率は 93% 程度とこの間には非常に大きな乖離が存在する。これはつまり、ゲノムが同一にも関わらず Pv11 細胞にはない蘇生率を劇的にあげるための転写制御関係をネムリユスリカ虫体は持っていると考えられ、このような転写制御関係を本研究で用いた方法論と同様に解析することで細胞においても、より強固な乾燥耐性機構の実現が可能になると考えられる。

本研究では 3.3.1 小節で明らかになった通り、再水和直後におけるミトコンドリア由来の転写物の量の比率が極端に増加する結果を得た。すなわち、再水和時特異的にミトコンドリアゲノム由来の RNA が高発現することで、死の淵から蘇ってくる Pv11 細胞がエネルギーを活発に産生させ始める機構が存在する可能性が見えてきたと言える。これは今までに発見されてきた生体障害を克服するという受動的な仕組みとは異なり、自ら生き返るためにエネルギーを生み出すという能動的な仕組みの発見といえる。この Pv11 細胞の再水和による蘇生過程でのミトコンドリアの関与及び関連遺伝子の同定、そしてそれら遺伝子の発現を駆動する転写因子を同定することで、死の淵からの脱却メカニズムが解明できれば、乾燥によって仮死状態として保存した生き物を自在に復活させることができるようになる可能性がある。

次の展望として、遺伝子制御ネットワーク推定に関する新規アルゴリズムの提案が考えられる。これまでの研究では転写因子とターゲット遺伝子は分けて制御関係を予測していたが、理想的には転写因子とターゲット遺伝子は分けずに全遺伝子を対象とした予測が行える必要がある。これにより転写因子の結合情報なしの解析が可能となる。しかし、これを行なうと対象とする遺伝子数が増加することとなる。ネムリユスリカの例では転写因子の数が 176 個に対して全遺伝子は 28415 個存在する。対象とする遺伝子数の増加は推定精度の減少に直結することが報告されている。そこで、このような遺伝子数増大に伴う推定精度の減少を抑えるために、回帰アルゴリズムの問題点に着目した。すなわち対象とする遺伝子数が少ない時にはデータへ程よく適合する結果を得られる一方で、対象とする遺伝子数が増大すると自由な回帰が可能になりデータに過度に適合してしまうことにより推

定精度が著しく落ちると考えられ、これを抑えることで対象とする遺伝子数増大に伴う推定精度の減少を抑えられると考えられる。そこでこのような過度なデータへの適合を回避するために推定時における制御関係に関する正則化の適用によってこれを実現することができると考えられる。

最後に乾燥耐性の制御による、乾燥しても死なない生命の創生に向けたシステム全体に対する包括的な展望を述べる。本研究では、Pv11細胞の乾燥耐性獲得期および再水和期に要する時間の観点から律速段階と考えられる遺伝子制御ネットワークに注目して、その制御ネットワークの推定を行なった。律速段階に注目することは現象の理解には有効であるが、その現象の制御を行なうためには律速段階以外の要素の欠如などを見逃すため不十分である。ここで、Pv11細胞の乾燥耐性獲得から再水和による細胞分裂再開までの過程を再度確認する。Pv11細胞の乾燥耐性獲得は、まず高濃度のトレハロースを含んだ培地処理により始まる。この時、細胞内部では、この通常の生理条件とは異なる状態を感知し細胞内に情報として伝えるシグナル伝達を行なうと考えられる。シグナル伝達により核内にこの情報が伝達すると今回推定した遺伝子制御機構が駆動され種々の乾燥耐性関連遺伝子の発現が誘導される。十分に乾燥耐性関連遺伝子の発現が準備されると、乾燥に対して多くの代謝活性を低下しながら無代謝状態へと移行する。そして再度水を与えられることでシグナル伝達を介した遺伝子発現が駆動され、代謝活性を元の状態に戻して細胞分裂を再開する。この一連の過程において、本研究で示唆されたことは遺伝子制御機構に限定されており、いまだにどのようなシグナル伝達経路を用いて遺伝子制御機構の駆動を行なうか、代謝活性の抑制と活性はどのように行なうかは明らかになっていない。Pv11細胞においては、これらの点を明らかにした研究はないが乾燥耐性を持つクマムシや線虫においては多くの研究がすでに行われてきた(1.2.1.2小々節参照)。これらの知見より、乾燥耐性においてPv11細胞が利用するシグナル伝達経路の候補としてはMAPK/ERK, JNK, p53およびAMPK経路が考えられる。また、本研究で推定した遺伝子制御ネットワークの最上流転写因子がNF- κ Bである、という知見を利用することでこれらシグナル伝達経路の中でNF- κ Bの発現に関わる経路を明らかにすれば乾燥耐性におけるシグナル伝達と遺伝子制御の間を繋ぐ関係を明らかにできる(図4.1)。次に、遺伝子制御と代謝の間を繋ぐ関係について議論する。トレハロースで覆われているPv11細胞においては、外界との物質のやりとりが遮断された状態になっており、Pv11細胞における代謝の活性および抑制は基質ではなく、これらの反応を触媒する酵素の発現を制御することで成り立っていると考えられる。これらの酵素をコードする遺伝子群は本研究で推定した遺伝子制御ネットワークの末端にあたるモジュールに含まれている。この代謝に密接に関わる酵素をコードする遺伝子群を含むモジュールと上流遺伝子制御に注目した解析を進めることで遺伝子制御と代謝の間を繋ぐ関係を明らかにできる(図4.1)。本研究で推定した遺伝子制御ネットワークに加えてシグナル伝達、代謝との関係性に基づき包括的な乾燥耐性システムを明らかにすることで、乾燥耐性を付与したい別生物種において欠損している要素(こ

れらは全てシグナル伝達経路を構成するタンパク質をコードする遺伝子，遺伝子制御ネットワークを構成する遺伝子，代謝反応を触媒するタンパク質をコードする遺伝子であるため)を遺伝子導入することによって乾燥しても死なない生物の創生が可能になる。

これらの展望を達成することにより，より包括的かつ正確な乾燥耐性制御機構の解明が行えると考えられる。

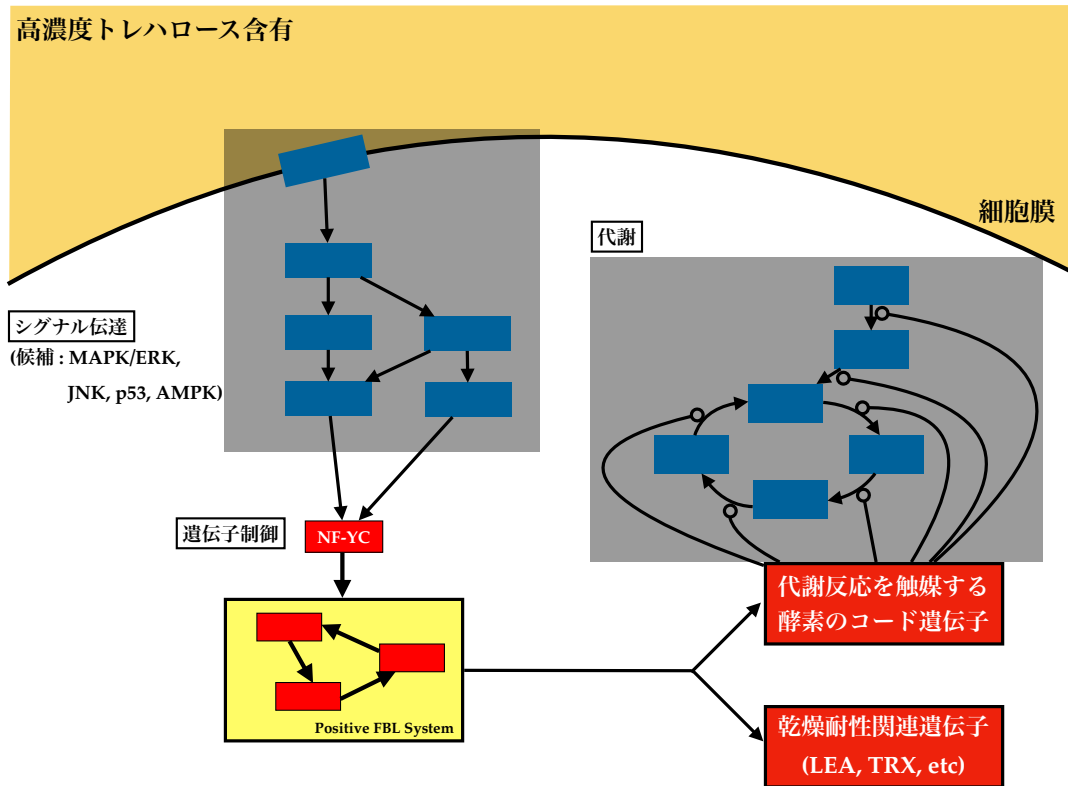


図 4.1: 乾燥耐性システムの全容と現在明らかにした制御機構

本研究では Pv11 細胞の乾燥耐性獲得期，再水和期に要する時間の観点から遺伝子制御に着目し最上流転写因子として NF-YC による Positive Feedback Loop 制御による乾燥耐性関連遺伝子群の遺伝子制御機構による乾燥耐性システムに対する知見を得た．一方で，この遺伝子制御機構の駆動を行なうシグナル伝達経路および下流に存在する代謝経路に関しては未知である．シグナル伝達についてはクマムシや線虫における乾燥耐性に関する先行研究により MAPK/ERK, JNK, p53, AMPK 経路が関わるということが報告されており，Pv11 細胞の乾燥耐性システムにおけるシグナル伝達としても最優先に検証すべき経路となる．また，本研究で得られた NF-YC が遺伝子制御機構における最上流転写因子であるという知見を利用することで，これらシグナル伝達経路のうち NF-YC の発現を誘導するシグナル伝達経路を見つければ，乾燥耐性システムにおけるシグナル伝達と遺伝子制御の関係を明らかにできる．続いて代謝に関しては，代謝活性および抑制を制御する反応の触媒である酵素は本研究で推定された遺伝子制御ネットワークにおいて最末端のモジュールに含まれている．これらのモジュール内の遺伝子群の発現変動をより詳細に解析することによって遺伝子制御と代謝の関係を明らかにできる．

謝辞

本研究は、著者が慶應義塾大学大学院理工学研究科後期博士課程在学中に、同大学理工学部生命情報学科 舟橋啓准教授の指導のもとに行なった研究の成果をまとめたものになります。

まず本研究をまとめるにあたり、常に親身に研究内容の相談に乗っていただき数えきれない助言を賜りました舟橋啓准教授に深謝いたします。舟橋啓准教授には本研究に関わる指導に限らず、学会発表におけるプレゼンテーション内容への指導や論文執筆時に際して重要な著作権に関わる指導など幅広い内容に関して昼夜を問わずに親身にご指導を賜りました。さらに定量生物学の会、ネムリユスリワークショップ、ロシアで開催された LIFE OF GENOMES など、本研究をまとめるにあたり非常に意義深い指摘を多数いただける発表の機会を多く与えていただきました。また、名誉ある同大学の助教という職への道を示していただき、これにより研究者であり教育者である大学の教員としての経験をさせていただきました。心より感謝申し上げます。

ご多用にも関わらず学位論文審査にて副査を引き受けてくださった慶應義塾大学理工学部生命情報学科 岡浩太郎教授、同大学理工学部生命情報学科 榊原康文教授、同大学理工学部生命情報学科 土居信英教授には本研究成果をまとめるにあたり、研究の正当性に留まらず本研究から得られた結果の解釈や意義について各専門分野の観点から大変貴重なご指摘を賜りました。この場をお借りして深く感謝いたします。

山陽小野田市立山口東京理科大学薬学部薬学科 広井賀子教授には研究を進めるにあたり生物学的な観点から多くのご指摘を賜りました。解析を中心とした研究遂行によりデータ解析の結果と生命現象の結びつきが薄くなりそうな時に、常にその関係を考える上で重要な知見の提供やご指導をいただきました。厚く御礼申し上げます。

農業・食品産業技術総合研究機構生物機能利用研究部門 黄川田隆洋博士には本研究を遂行するにあたり必要不可欠な Pv11 細胞のトランスクリプトームデータを提供いただきました。黄川田隆洋博士には研究を進めるにあたり解析結果に関するディスカッションを常に快く承諾いただき、乾燥耐性を軸とした生理学的な観点から学術論文をまとめる上で重要なご指摘をいただきました。また黄川田研究室の研究員、学生の方々には研究をまとめるにあたり議論をさせていただき、多くの重要なご指摘をいただきました。黄川田隆洋博士ならびに生物機能利用研究部門の Richard Cornette 博士、宮田佑吾博士、十亀陽一郎博士 (現 福島工業高等

専門学校 特命助教), 東京大学大学院新領域科学研究科先端生命専攻の徳本翔子女史には深く感謝いたします。

農業生物資源研究所 (現 農業・食品産業技術総合研究機構) 末次克行博士 (故人) には研究を開始した当初にネムリユスリカゲノムデータの取り扱い方法についてご指導いただきました。初歩的な質問にも迅速にご指導を賜りましたこと、この場をお借りして感謝申し上げます。

理化学研究所 KFU 応用ゲノム特別ユニット Oleg Gusev 博士には本研究を進めるにあたりゲノムデータの整備や本研究をロシアで開催された LIFE OF GENOMES で招待講演として発表する機会をいただきました。この場をお借りして厚く御礼申し上げます。

Kazan Federal University の Elena Shagimardanova 博士には本研究におけるデータ取得に関してご協力をいただきました。また同大学の Alexander Nesmelov 博士, Ruslan Deviatiiarov 君には乾燥耐性関連遺伝子の同定に際して多数のご助言をいただきました。この場をお借りして感謝申し上げます。

東京工業大学物質理工学院の倉科佑太助教には広井賀子教授と共に山陽小野田市立山口東京理科大学で開催した CELLab-SOCU Summer Symposium の際にご一緒させていただき、本研究の内容を議論する機会をいただきました。また、慶應義塾大学理工学部機械工学科, システムデザイン工学科の有志の学生発信の研究会である細胞勉強会 (現 CELLab) にお呼びいただき、生物学とは異なる分野で研究をされている学生や研究者の方々と交流する機会を与えていただきました。また、同勉強会に参加していた開放環境科学専攻の今城哉裕博士, 総合デザイン工学専攻の渡邊應文君とは同期ということもあり活発に互いの研究に関して議論させていただきました。誠にありがとうございました。

著者が所属するシステム生物学研究室の OB である平岩巧博士には研究を進める上で大事なことや日常生活の中で多くのディスカッションをする機会をいただきました。本当にありがとうございました。また同じ後期博士課程の徳岡雄大君とは互いの研究について指摘し合い、切磋琢磨して研究を進めることができる心強い仲間でした。本当に感謝いたします。また、前期博士課程の比企佑介君は、システム生物学研究室で私以外で初めてネムリユスリカを対象とした研究を進めてくれて乾燥耐性研究の広がりを作ってくれました。本当にありがとうございました。他にもシステム生物学研究室の OB・OG 含め皆様には多くの協力をいただきました。皆様には本当に感謝申し上げます。

後期博士課程では研究のすゝめ奨学金による支援を受けました。この奨学金のおかげで学費の心配をせずに研究に集中することができました。この場をお借りして感謝申し上げます。

最後に両親への感謝を述べたいと思います。研究を進める中でどうしても連日遅くまでかかることが多く不規則な生活にどうしてもならざるを得ない状況の中でも健康に研究のことにのみ専心することができたのは偏に父母の支えがあってこそでした。経済的な面のみではなく、悩みごとの相談に乗ってくれたり毎日夜分遅くに帰っても夕食の準備をしてくれた精神的な支えのおかげで、本研究をここまでまとめることができました。両親には感謝してもしきれません。また、何気ない会話で心を和ませてくれた妹にも感謝いたします。今まで本当にありがとうございました。

参考文献

- [1] John H Crowe, Folkert A Hoekstra, and Lois M Crowe. Anhydrobiosis. *Annu. Rev. Physiol.*, Vol. 54, No. 1, pp. 579–599, 1992.
- [2] Jens Lapinski and Alan Tunnacliffe. Anhydrobiosis without trehalose in bdelloid rotifers. *FEBS Lett.*, Vol. 553, No. 3, pp. 387–390, 2003.
- [3] Weronika Wełnicz, Markus A Grohme, Łukasz Kaczmarek, Ralph O Schill, and Marcus Frohme. Anhydrobiosis in tardigrades—the last decade. *J. Insect Physiol.*, Vol. 57, No. 5, pp. 577–583, 2011.
- [4] Robert McSorley. Adaptations of nematodes to environmental extremes. *Fla. Entomol.*, Vol. 86, No. 2, pp. 138–143, 2003.
- [5] Folkert A Hoekstra, Elena A Golovina, and Julia Buitink. Mechanisms of plant desiccation tolerance. *Trends Plant Sci.*, Vol. 6, No. 9, pp. 431–438, 2001.
- [6] Takahiro Kikawada, Noboru Minakawa, Masahiko Watanabe, and Takashi Okuda. Factors inducing successful anhydrobiosis in the African chironomid *Polypedilum vanderplanki*: significance of the larval tubular nest. *Integr. Comp. Biol.*, Vol. 45, No. 5, pp. 710–714, 2005.
- [7] Yuichi Nakahara, Shigeo Imanishi, Kanako Mitsumasu, Yasushi Kanamori, Ken-ichi Iwata, Masahiko Watanabe, Takahiro Kikawada, and Takashi Okuda. Cells from an anhydrobiotic chironomid survive almost complete desiccation. *Cryobiology*, Vol. 60, No. 2, pp. 138–146, 2010.
- [8] Van Leeuwenhoek, A. (1702) Letter to Hendrik van Bleyswijk dated 9 February 1702 (in Italian).
- [9] クリフォードドーベル. レーベンフックの手紙, p. 263. 九州大学出版会, 12 2003.
- [10] Masahiko Watanabe. Anhydrobiosis in invertebrates. *Appl. Entomol. Zool.*, Vol. 41, No. 1, pp. 15–31, 2006.

- [11] D Keilin. The problem of anabiosis or latent life: history and current concept. the Leeuwenhoek lecture. *Proc R Soc Lond B*, Vol. 150, No. 939, pp. 148–191, 1959.
- [12] John H Crowe. Evaporative water loss by tardigrades under controlled relative humidities. *Biol. Bull.*, Vol. 142, No. 3, pp. 407–416, 1972.
- [13] Paul H Yancey, Mary E Clark, Steven C Hand, R David Bowlus, and George N Somero. Living with water stress: evolution of osmolyte systems. *Science*, Vol. 217, No. 4566, pp. 1214–1222, 1982.
- [14] John H Crowe, James S Clegg, and Lois M Crowe. *Anhydrobiosis: the water replacement hypothesis*, pp. 440–455. Springer, 1998.
- [15] Roberto Bertolani, Roberto Guidetti, Ingemar K Jönsson, Tiziana Altiero, Deborah Boschini, and Lorena Rebecchi. Experiences with dormancy in tardigrades. *J Limnol*, Vol. 63, No. 1s, pp. 16–25, 2004.
- [16] Selvan Bavan, Volko A Straub, Mark L Blaxter, and Steven J Ennion. A P2X receptor from the tardigrade species *Hypsibius dujardini* with fast kinetics and sensitivity to zinc and copper. *BMC Evol. Biol.*, Vol. 9, No. 17, 2009.
- [17] Lois M Crowe, John H Crowe, Alan Rudolph, Christopher Womersley, and Lori Appel. Preservation of freeze-dried liposomes by trehalose. *Arch. Biochem. Biophys*, Vol. 242, No. 1, pp. 240–247, 1985.
- [18] Kazuyo Watanabe, Shigeo Imanishi, Gaku Akiduki, Richard Cornette, and Takashi Okuda. Air-dried cells from the anhydrobiotic insect, *Polypedilum vanderplanki*, can survive long term preservation at room temperature and retain proliferation potential after rehydration. *Cryobiology*, Vol. 73, No. 1, pp. 93–98, 2016.
- [19] Oleg Gusev, Yoshitaka Suetsugu, Richard Cornette, Takeshi Kawashima, Maria D Logacheva, Alexey S Kondrashov, Aleksey A Penin, Rie Hatanaka, Shingo Kikuta, Sachiko Shimura, et al. Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. *Nat Commun*, Vol. 5, No. 4784, 2014.
- [20] Richard Cornette, Oleg Gusev, and Takahiro Kikawada. New insights about anhydrobiosis in *Polypedilum vanderplanki* (Diptera, chironomidae) from metabolome analysis. *Cryobiology*, Vol. 85, No. 137, 2018.

- [21] Elham Schokraie, Uwe Warnken, Agnes Hotz-Wagenblatt, Markus A Grohme, Steffen Hengherr, Frank Förster, Ralph O Schill, Marcus Frohme, Thomas Dandekar, and Martina Schnölzer. Comparative proteome analysis of *Milnesium tardigradum* in early embryonic state versus adults in active and anhydrobiotic state. *PLOS ONE*, Vol. 7, No. 9, p. e45682, 2012.
- [22] Zebo Huang, Matthew C Banton, and Alan Tunnacliffe. Modeling anhydrobiosis: activation of the mitogen-activated protein kinase ERK by dehydration in both human cells and nematodes. *J Exp Zool A Ecol Genet Physiol*, Vol. 313, No. 10, pp. 660–670, 2010.
- [23] Koyuki Kondo, Masaru Mori, Masaru Tomita, and Kazuharu Arakawa. AMPK activity is required for the induction of anhydrobiosis in a tardigrade *Hypsibius exemplaris*, and its potential up-regulator is PP2A. *Genes to Cells*, Vol. 24, No. 12, pp. 768–780, 2019.
- [24] Xiangyu Zhou, Isao Naguro, Hidenori Ichijo, and Kengo Watanabe. Mitogen-activated protein kinases as key players in osmotic stress signaling. *Biochim Biophys Acta Gen Subj*, Vol. 1860, No. 9, pp. 2037–2052, 2016.
- [25] Cihan Erkut, Andrej Vasilj, Sebastian Boland, Bianca Habermann, Andrej Shevchenko, and Teymuras V Kurzchalia. Molecular strategies of the *Caenorhabditis elegans* dauer larva to survive extreme desiccation. *PLOS ONE*, Vol. 8, No. 12, p. e82473, 2013.
- [26] Lei Zhang, Akinori Ohta, Masamichi Takagi, and Ryozo Imai. Expression of plant group 2 and group 3 *lea* genes in *Saccharomyces cerevisiae* revealed functional divergence among LEA proteins. *J. Biochem.*, Vol. 127, No. 4, pp. 611–616, 2000.
- [27] Kshamata Goyal, Claudia Pinelli, Sarah L Maslen, Rakesh K Rastogi, Elaine Stephens, and Alan Tunnacliffe. Dehydration-regulated processing of late embryogenesis abundant protein in a desiccation-tolerant nematode. *FEBS Lett.*, Vol. 579, No. 19, pp. 4093–4098, 2005.
- [28] Jose L Soulages, Kangmin Kim, Christina Walters, and John C Cushman. Temperature-induced extended helix/random coil transitions in a group 1 late embryogenesis-abundant protein from soybean. *Plant Physiology*, Vol. 128, No. 3, pp. 822–832, 2002.
- [29] Rie Hatanaka, Yuka Hagiwara-Komoda, Takao Furuki, Yasushi Kanamori, Mika Fujita, Richard Cornette, Minoru Sakurai, Takashi Okuda, and

- Takahiro Kikawada. An abundant LEA protein in the anhydrobiotic midge, PvLEA4, acts as a molecular shield by limiting growth of aggregating protein particles. *Insect Biochem. Mol. Biol.*, Vol. 43, No. 11, pp. 1055–1067, 2013.
- [30] Yuki Yoshida, Georgios Koutsovoulos, Dominik R Laetsch, Lewis Stevens, Sujai Kumar, Daiki D Horikawa, Kyoko Ishino, Shiori Komine, Takekazu Kunieda, Masaru Tomita, et al. Comparative genomics of the tardigrades *Hypsibius dujardini* and *Ramazzottius varieornatus*. *PLoS Biol.*, Vol. 15, No. 7, p. e2002266, 2017.
- [31] Fredrik Åslund, Kurt D Berndt, and Arne Holmgren. Redox potentials of glutaredoxins and other thiol-disulfide oxidoreductases of the thioredoxin superfamily determined by direct protein-protein redox equilibria. *J. Biol. Chem.*, Vol. 272, No. 49, pp. 30780–30786, 1997.
- [32] Wendell Q Sun and A Carl Leopold. Cytoplasmic vitrification and survival of anhydrobiotic organisms. *Comp. Biochem. Physiol. A Physiol.*, Vol. 117, No. 3, pp. 327–333, 1997.
- [33] Yoichiro Sogame and Takahiro Kikawada. Current findings on the molecular mechanisms underlying anhydrobiosis in *Polypedilum vanderplanki*. *Curr Opin Insect Sci*, Vol. 19, pp. 16–21, 2017.
- [34] Shilpi Khare, Carole L Linster, and Steven G Clarke. The interplay between protein L-isoaspartyl methyltransferase activity and insulin-like signaling to extend lifespan in *Caenorhabditis elegans*. *PLOS ONE*, Vol. 6, No. 6, p. e20850, 2011.
- [35] Thorsten Burmester and Thomas Hankeln. The respiratory proteins of insects. *J. Insect Physiol.*, Vol. 53, No. 4, pp. 285–294, 2007.
- [36] Shin G Goto, Richard E Lee Jr, and David L Denlinger. Aquaporins in the Antarctic midge, an extremophile that relies on dehydration for cold survival. *Biol. Bull*, Vol. 229, No. 1, pp. 47–57, 2015.
- [37] Peter Agre and David Kozono. Aquaporin water channels: molecular mechanisms for human diseases1. *FEBS Lett.*, Vol. 555, No. 1, pp. 72–78, 2003.
- [38] Oscar JM Goddijn and Kees van Dun. Trehalose metabolism in plants. *Trends Plant Sci.*, Vol. 4, No. 8, pp. 315–319, 1999.

- [39] Pavel V Mazin, Elena Shagimardanova, Olga Kozlova, Alexander Cherkasov, Roman Sutormin, Vita V Stepanova, Alexey Stupnikov, Maria Logacheva, Aleksey Penin, Yoichiro Sogame, et al. Cooption of heat shock regulatory system for anhydrobiosis in the sleeping chironomid *Polypedilum vanderplanki*. *Proc. Natl. Acad. Sci. USA*, Vol. 115, No. 10, pp. E2477–E2486, 2018.
- [40] Andras Orosz, Jan Wisniewski, and Carl Wu. Regulation of *Drosophila* heat shock factor trimerization: global sequence requirements and independence of nuclear localization. Vol. 16, No. 12, pp. 7018–7030, 1996.
- [41] 堀川大樹. ヨコヅナクマムシ：極限環境動物のモデルとして. 極限環境微生物学会誌, Vol. 7.2, No. 1, pp. 25–28, 2008.
- [42] Christopher Womersley and Cherie Ching. Natural dehydration regimes as a prerequisite for the successful induction of anhydrobiosis in the nematode *Rotylenchulus reniformis*. *J. Exp. Biol.*, Vol. 143, No. 1, pp. 359–372, 1989.
- [43] Manuela Caprioli and Claudia Ricci. Recipes for successful anhydrobiosis in bdelloid rotifers. In *Rotifera IX*, pp. 13–17. Springer, 2001.
- [44] Koyuki Kondo, Takeo Kubo, and Takekazu Kunieda. Suggested involvement of PP1/PP2A activity and de novo gene expression in anhydrobiotic survival in a tardigrade, *Hypsibius dujardini*, by chemical genetic approach. *PLOS ONE*, Vol. 10, No. 12, p. e0144803, 2015.
- [45] Emanuel Gonçalves, Joachim Bucher, Anke Ryll, Jens Niklas, Klaus Mauch, Steffen Klamt, Miguel Rocha, and Julio Saez-Rodriguez. Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models. *Mol. Biosyst.*, Vol. 9, No. 7, pp. 1576–1583, 2013.
- [46] Jason A Papin, Tony Hunter, Bernhard O Palsson, and Shankar Subramaniam. Reconstruction of cellular signalling networks and analysis of their properties. *Nat. Rev. Mol. Cell Biol.*, Vol. 6, pp. 99–111, 2005.
- [47] Christine N Tennyson, Henry J Klamut, and Ronald G Worton. The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat. Genet.*, Vol. 9, pp. 184–190, 1995.
- [48] David A Wharton. *16. Nematode Survival Strategies*. CRC Press, 2002.
- [49] Leena J Thorat, Sushama M Gaikwad, and Bimalendu B Nath. Trehalose as an indicator of desiccation stress in *Drosophila melanogaster* larvae: a

- potential marker of anhydrobiosis. *Biochem. Biophys. Res. Commun.*, Vol. 419, No. 4, pp. 638–642, 2012.
- [50] Andreas PM Weber, Robin J Horst, Guillaume G Barbier, and Christine Oesterhelt. Metabolism and metabolomics of eukaryotes living under extreme conditions. *Int. Rev. Cytol.*, Vol. 256, pp. 1–34, 2007.
- [51] Dinakar Challabathula and Dorothea Bartels. Desiccation tolerance in resurrection plants: new insights from transcriptome, proteome and metabolome analysis. *Front. Plant Sci.*, Vol. 4, No. 482, 2013.
- [52] Matthew C Banton and Alan Tunnacliffe. MAPK phosphorylation is implicated in the adaptation to desiccation stress in nematodes. *J. Exp. Biol.*, Vol. 215, No. 24, pp. 4288–4298, 2012.
- [53] Masahiko Watanabe, Yuichi Nakahara, Tetsuya Sakashita, Takahiro Kikawada, Akihiko Fujita, Nobuyuki Hamada, Daiki D Horikawa, Seiichi Wada, Yasuhiko Kobayashi, and Takashi Okuda. Physiological changes leading to anhydrobiosis improve radiation tolerance in *Polypedilum vanderplanki* larvae. *J. Insect Physiol.*, Vol. 53, No. 6, pp. 573–579, 2007.
- [54] Mauro B França, Anita D Panek, and Elis CA Eleutherio. Oxidative stress and its effects during dehydration. *Comp. Biochem. Physiol., Part A Mol. Integr. Physiol.*, Vol. 146, No. 4, pp. 621–631, 2007.
- [55] Daiki D Horikawa and Seigo Higashi. Desiccation tolerance of the tardigrade *Milnesium tardigradum* collected in Sapporo, Japan, and Bogor, Indonesia. *Zool. Sci.*, Vol. 21, No. 8, pp. 813–817, 2004.
- [56] Eric H Davidson. *The regulatory genome: gene regulatory networks in development and evolution*. Elsevier, 2010.
- [57] Shubhada R Kulkarni, Dries Vanechoutte, Jan Van de Velde, and Klaas Vandepoele. TF2Network: predicting transcription factor regulators and gene regulatory networks in *Arabidopsis* using publicly available binding site information. *Nucleic Acids Res.*, Vol. 46, No. 6, pp. e31–e31, 2017.
- [58] Shane Neph, Andrew B Stergachis, Alex Reynolds, Richard Sandstrom, Elhanan Borenstein, and John A Stamatoyannopoulos. Circuitry and dynamics of human transcription factor regulatory networks. *Cell*, Vol. 150, No. 6, pp. 1274–1286, 2012.

- [59] Araceli M Huerta, Heladia Salgado, Denis Thieffry, and Julio Collado-Vides. RegulonDB: a database on transcriptional regulation in *Escherichia coli*. *Nucleic Acids Res.*, Vol. 26, No. 1, pp. 55–59, 1998.
- [60] Tong Ihn Lee, Nicola J Rinaldi, François Robert, Duncan T Odom, Ziv Bar-Joseph, Georg K Gerber, Nancy M Hannett, Christopher T Harbison, Craig M Thompson, Itamar Simon, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, Vol. 298, No. 5594, pp. 799–804, 2002.
- [61] Sushmita Roy, Jason Ernst, Peter V Kharchenko, Pouya Kheradpour, Nicolas Negre, Matthew L Eaton, Jane M Landolin, Christopher A Bristow, Lijia Ma, Michael F Lin, et al. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science*, Vol. 330, No. 6012, pp. 1787–1797, 2010.
- [62] Dijun Chen, Wenhao Yan, Liang-Yu Fu, and Kerstin Kaufmann. Architecture of gene regulatory networks controlling flower development in *Arabidopsis thaliana*. *Nat Commun*, Vol. 9, No. 4534, 2018.
- [63] Nor IA Rahman, Nor A Abdul Murad, Mohammad M Mollah, Rahman Jamal, and Roslan Harun. NFIX as a master regulator for lung cancer progression. *Front Pharmacol*, Vol. 8, No. 540, 2017.
- [64] Nabil Guelzim, Samuele Bottani, Paul Bourguine, and François Képès. Topological and causal structure of the yeast transcriptional regulatory network. *Nat. Genet.*, Vol. 31, pp. 60–63, 2002.
- [65] Reka Albert. Scale-free networks in cell biology. *J. Cell. Sci.*, Vol. 118, No. 21, pp. 4947–4957, 2005.
- [66] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, Vol. 298, No. 5594, pp. 824–827, 2002.
- [67] Shai S Shen-Orr, Ron Milo, Shmoolik Mangan, and Uri Alon. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.*, Vol. 31, pp. 64–68, 2002.
- [68] Jing Yu, V Anne Smith, Paul P Wang, Alexander J Hartemink, and Erich D Jarvis. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics*, Vol. 20, No. 18, pp. 3594–3603, 2004.

- [69] Pietro Zoppoli, Sandro Morganello, and Michele Ceccarelli. TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinformatics*, Vol. 11, No. 154, 2010.
- [70] Alexandre Irrthum, Louis Wehenkel, Pierre Geurts, et al. Inferring regulatory networks from expression data using tree-based methods. *PLOS ONE*, Vol. 5, No. 9, p. e12776, 2010.
- [71] Sungjoon Park, Jung Min Kim, Wonho Shin, Sung Won Han, Minji Jeon, Hyun Jin Jang, Ik-Soon Jang, and Jaewoo Kang. BTNET: boosted tree based gene regulatory network inference algorithm using time-course measurement data. *BMC Syst. Biol.*, Vol. 12, No. 20, 2018.
- [72] Anne-Claire Haury, Fantine Mordelet, Paola Vera-Licona, and Jean-Philippe Vert. TIGRESS: trustful inference of gene regulation using stability selection. *BMC Syst. Biol.*, Vol. 6, No. 145, 2012.
- [73] Robert Küffner, Tobias Petri, Pegah Tavakkolkhah, Lukas Windhager, and Ralf Zimmer. Inferring gene regulatory networks by ANOVA. *Bioinformatics*, Vol. 28, No. 10, pp. 1376–1382, 2012.
- [74] Vân Anh Huynh-Thu and Pierre Geurts. dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. *Sci. Rep*, Vol. 8, No. 1, p. 3384, 2018.
- [75] Vân Anh Huynh-Thu and Guido Sanguinetti. Combining tree-based and dynamical systems for the inference of gene regulatory networks. *Bioinformatics*, Vol. 31, No. 10, pp. 1614–1622, 2015.
- [76] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Ann. Stat.*, pp. 1189–1232, 2001.
- [77] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, Vol. 37, No. 3, pp. 424–438, 1969.
- [78] Peter J Park. ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.*, Vol. 10, pp. 669–680, 2009.
- [79] Rasika Mundade, Hatice Gulcin Ozer, Han Wei, Lakshmi Prabhu, and Tao Lu. Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle*, Vol. 13, No. 18, pp. 2847–2852, 2014.

- [80] Alexander J. Stewart, Sridhar Hannenhalli, and Joshua B. Plotkin. Why Transcription factor binding sites are ten nucleotides long. *Genetics*, Vol. 192, No. 3, pp. 973–985, 2012.
- [81] Philip Machanick and Timothy L. Bailey. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*, Vol. 27, No. 12, pp. 1696–1697, 2011.
- [82] Aziz Khan, Oriol Fornes, Arnaud Stigliani, Marius Gheorghe, Jaime A Castro-Mondragon, Robin van der Lee, Adrien Bessy, Jeanne Cheneby, Shubhada R Kulkarni, Ge Tan, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.*, Vol. 46, No. D1, pp. D260–D266, 2017.
- [83] Lihua Julie Zhu, Ryan G Christensen, Majid Kazemian, Christopher J Hull, Metewo Selase Enuameh, Matthew D Basciotta, Jessie A Brasefield, Cong Zhu, Yuna Asriyan, David S Lapointe, et al. FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res.*, Vol. 39, pp. D111–D117, 2010.
- [84] Matthew T Weirauch, Ally Yang, Mihai Albu, Atina G Cote, Alejandro Montenegro-Montero, Philipp Drewe, Hamed S Najafabadi, Samuel A Lambert, Ishminder Mann, Kate Cook, et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*, Vol. 158, No. 6, pp. 1431–1443, 2014.
- [85] Robert C McLeay and Timothy L Bailey. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics*, Vol. 11, No. 165, 2010.
- [86] Martin C Frith, Yutao Fu, Liqun Yu, Jiang-Fan Chen, Ulla Hansen, and Zhiping Weng. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res.*, Vol. 32, No. 4, pp. 1372–1381, 2004.
- [87] Dimos Gaidatzis, Anita Lerch, Florian Hahne, and Michael B Stadler. QuasR: quantification and annotation of short reads in R. *Bioinformatics*, Vol. 31, No. 7, pp. 1130–1132, 2015.
- [88] Jianqiang Sun, Tomoaki Nishiyama, Kentaro Shimizu, and Koji Kadota. TCC: an R package for comparing tag count data with robust normalization strategies. *BMC Bioinformatics*, Vol. 14, No. 219, 2013.

- [89] Mark D Robinson, Davis J McCarthy, and Gordon K Smyth. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, Vol. 26, No. 1, pp. 139–140, 2010.
- [90] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol.*, Vol. 57, No. 1, pp. 289–300, 1995.
- [91] Koji Kadota, Tomoaki Nishiyama, and Kentaro Shimizu. A normalization strategy for comparing tag count data. *Algorithms Mol Biol.*, Vol. 7, No. 5, 2012.
- [92] Gregory R. Warnes, Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz, and Bill Venables. *gplots: Various R Programming Tools for Plotting Data*, 2015. R package version 2.17.0.
- [93] Ana Conesa, Stefan Götz, Juan Miguel García-Gómez, Javier Terol, Manuel Talón, and Montserrat Robles. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, Vol. 21, No. 18, pp. 3674–3676, 2005.
- [94] Fran Supek, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE.*, Vol. 6, No. 7, p. e21800, 2011.
- [95] Andreas Schlicker, Francisco S Domingues, Jörg Rahnenführer, and Thomas Lengauer. A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics*, Vol. 7, No. 302, 2006.
- [96] Aron Marchler-Bauer, Shennan Lu, John B Anderson, Farideh Chitsaz, Myra K Derbyshire, Carol DeWeese-Scott, Jessica H Fong, Lewis Y Geer, Renata C Geer, Noreen R Gonzales, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.*, Vol. 39, pp. D225–D229, 2010.
- [97] Guy Yachdav, Sebastian Wilzbach, Benedikt Rauscher, Robert Sheridan, Ian Sillitoe, James Procter, Suzanna E Lewis, Burkhard Rost, and Tatyana Goldberg. MSASviewer: interactive JavaScript visualization of multiple sequence alignments. *Bioinformatics*, Vol. 32, No. 22, pp. 3501–3503, 2016.
- [98] Takahiro G Yamada, Yoshitaka Suetsugu, Ruslan Deviatiiarov, Oleg Gusev, Richard Cornette, Alexander Nesmelov, Noriko Hiroi, Takahiro Kikawada,

- and Akira Funahashi. Transcriptome analysis of the anhydrobiotic cell line Pv11 infers the mechanism of desiccation tolerance and recovery. *Sci Rep*, Vol. 8, No. 17941, 2018.
- [99] Thomas Schell, Andreas E Kulozik, and Matthias W Hentze. Integration of splicing, transport and translation to achieve mRNA quality control by the nonsense-mediated decay pathway. *Genome Biol.*, Vol. 3, No. 3, pp. reviews1006.1–reviews1006.6, 2002.
- [100] Yusong Yang, Ji-Zhong Cheng, Sharad S Singhal, Manjit Saini, Utpal Pandya, Sanjay Awasthi, and Yogesh C Awasthi. Role of glutathione S-transferases in protection against lipid peroxidation overexpression of HGSTA2-2 in K562 cells protects against hydrogen peroxide-induced apoptosis and inhibits JNK and caspase 3 activation. *J. Biol. Chem.*, Vol. 276, No. 22, pp. 19220–19230, 2001.
- [101] Motoaki Seki, Mari Narusaka, Junko Ishida, Tokihiko Nanjo, Miki Fujita, Youko Oono, Asako Kamiya, Maiko Nakajima, Akiko Enju, Tetsuya Sakurai, et al. Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.*, Vol. 31, No. 3, pp. 279–292, 2002.
- [102] B Burchell, CH Brierley, G Monaghan, and DJ Clarke. The structure and function of the UDP-glucuronosyltransferase gene family. In *Adv Pharmacol.*, Vol. 42, pp. 335–338. Elsevier, 1997.
- [103] Kenneth Bachmann. *Drug metabolism*. Elsevier, 2009.
- [104] Cyril F Bourgeois, Franck Mortreux, and Didier Auboeuf. The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nat. Rev. Mol. Cell Biol.*, Vol. 17, pp. 426–438, 2016.
- [105] Zhongyu Xie, Junbiao Dai, Lunzhi Dai, Minjia Tan, Zhongyi Cheng, Yeming Wu, Jef D Boeke, and Yingming Zhao. Lysine succinylation and lysine malonylation in histones. *Mol. Cell Proteomics.*, Vol. 11, No. 5, pp. 100–107, 2012.
- [106] Sami N Guzder, Patrick Sung, Louise Prakash, and Satya Prakash. Yeast Rad7-Rad16 complex, specific for the nucleotide excision repair of the non-transcribed DNA strand, is an ATP-dependent DNA damage sensor. *J. Biol. Chem.*, Vol. 272, No. 35, pp. 21665–21668, 1997.

- [107] Oriana Muzzin, Elizabeth A Campbell, Lulin Xia, Elena Severinova, Seth A Darst, and Konstantin Severinov. Disruption of *Escherichia coli* hepA, an RNA polymerase-associated protein, causes UV sensitivity. *J. Biol. Chem.*, Vol. 273, No. 24, pp. 15157–15161, 1998.
- [108] Richard D Wood, Michael Mitchell, John Sgouros, and Tomas Lindahl. Human DNA repair genes. *Science.*, Vol. 291, No. 5507, pp. 1284–1289, 2001.
- [109] Richard D Wood, Michael Mitchell, and Tomas Lindahl. Human DNA repair genes, 2005. *Mutat. Res.*, Vol. 577, No. 1, pp. 275–283, 2005.
- [110] Masahiko Watanabe, Takahiro Kikawada, and Takashi Okuda. Increase of internal ion concentration triggers trehalose synthesis associated with cryptobiosis in larvae of *Polypedilum vanderplanki*. *J. Exp. Biol.*, Vol. 206, No. 13, pp. 2281–2286, 2003.
- [111] M Watanabe, T Kikawada, N Minagawa, F Yukuhiro, and T Okuda. Mechanism allowing an insect to survive complete dehydration and extreme temperatures. *J. Exp. Biol.*, Vol. 205, No. 18, pp. 2799–2802, 2002.
- [112] Takahiro Kikawada, Yuichi Nakahara, Yasushi Kanamori, Ken-ichi Iwata, Masahiko Watanabe, Brian McGee, Alan Tunnacliffe, and Takashi Okuda. Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. *Biochem. Biophys. Res. Commun.*, Vol. 348, No. 1, pp. 56–61, 2006.
- [113] Takahiro Kikawada, Ayako Saito, Yasushi Kanamori, Mika Fujita, Katarzyna Śnigórska, Masahiko Watanabe, and Takashi Okuda. Dehydration-inducible changes in expression of two aquaporins in the sleeping chironomid, polypedilum vanderplanki. *Biochim Biophys Acta Biomembr*, Vol. 1778, No. 2, pp. 514–520, 2008.
- [114] Eric Weterings and David J Chen. The endless tale of non-homologous end-joining. *Cell Res.*, Vol. 18, pp. 114–124, 2008.
- [115] Anthony J Davis and David J Chen. DNA double strand break repair via non-homologous end-joining. *Transl Cancer Res.*, Vol. 2, No. 3, pp. 130–143, 2013.
- [116] Zhiyong Mao, Michael Bozzella, Andrei Seluanov, and Vera Gorbunova. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst.)*, Vol. 7, No. 10, pp. 1765–1771, 2008.

- [117] Sebastien Dupont, Alexander Rapoport, Patrick Gervais, and Laurent Beney. Survival kit of *Saccharomyces cerevisiae* for anhydrobiosis. *Appl. Microbiol. Biotechnol.*, Vol. 98, No. 21, pp. 8821–8834, 2014.
- [118] James S Clegg. Cryptobiosis—a peculiar state of biological organization. *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.*, Vol. 128, No. 4, pp. 613–624, 2001.
- [119] Rie Hatanaka, Oleg Gusev, Richard Cornette, Sachiko Shimura, Shingo Kikuta, Jun Okada, Takashi Okuda, and Takahiro Kikawada. Diversity of the expression profiles of late embryogenesis abundant (LEA) protein encoding genes in the anhydrobiotic midge *Polypedilum vanderplanki*. *Planta*, Vol. 242, No. 2, pp. 451–459, 2015.
- [120] Richard Cornette, Yasushi Kanamori, Masahiko Watanabe, Yuichi Nakahara, Oleg Gusev, Kanako Mitsumasu, Keiko Kadono-Okuda, Michihiko Shimomura, Kazuei Mita, Takahiro Kikawada, et al. Identification of anhydrobiosis-related genes from an expressed sequence tag database in the cryptobiotic midge *Polypedilum vanderplanki* (diptera; chironomidae). *J. Biol. Chem.*, Vol. 285, No. 46, pp. 35889–35899, 2010.
- [121] Simon Andrews. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, 2010.
- [122] Chien-Chi Lo and Patrick SG Chain. Rapid evaluation and quality control of next generation sequencing data with FaQCs. *BMC Bioinformatics*, Vol. 15, No. 366, 2014.
- [123] Daehwan Kim, Ben Langmead, and Steven L Salzberg. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, Vol. 12, pp. 357–360, 2015.
- [124] Uwe Ligges and Martin Mächler. Scatterplot3d - an R package for visualizing multivariate data. *J Stat Softw*, Vol. 8, No. 11, pp. 1–20, 2003.
- [125] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, Vol. 15, No. 550, 2014.
- [126] Sarah K Kummerfeld and Sarah A Teichmann. DBD: a transcription factor prediction database. *Nucleic Acids Res.*, Vol. 34, pp. D74–D81, 2006.

- [127] Hanbo Chen. VennDiagram: Generate high-resolution Venn and Euler plots, 2018. R package version 1.6.20.
- [128] Maximino Aldana, Enrique Balleza, Stuart Kauffman, and Osbaldo Resendiz. Robustness and evolvability in genetic regulatory networks. *J. Theor. Biol.*, Vol. 245, No. 3, pp. 433–448, 2007.
- [129] Roland Wiese, Markus Eiglsperger, and Michael Kaufmann. yFiles—visualization and automatic layout of graphs. In *Graph Drawing Software*, pp. 173–191. Springer, 2004.
- [130] Gabor Csardi, Tamas Nepusz, et al. The igraph software package for complex network research. *InterJournal, Complex Systems*, Vol. 1695, No. 5, pp. 1–9, 2006.
- [131] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, Vol. 5, No. 1, pp. 17–60, 1960.
- [132] Peter Langfelder and Steve Horvath. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, Vol. 9, No. 559, 2008.
- [133] Andy M Yip and Steve Horvath. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics*, Vol. 8, No. 22, 2007.
- [134] Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Commun Stat Theory Methods*, Vol. 3, No. 1, pp. 1–27, 1974.
- [135] Seth Falcon and Gentleman Robert. Using GOstats to test gene lists for GO term association. *Bioinformatics*, Vol. 23, No. 2, pp. 257–8, 2007.
- [136] Paul Shannon and Matt Richards. *MotifDb: An Annotated Collection of Protein-DNA Binding Sequence Motifs*, 2018. R package version 1.24.1.
- [137] Bernhard Pfaff. VAR, SVAR and SVEC models: Implementation within R package vars. *J Stat Softw*, Vol. 27, No. 4, pp. 1–32, 2008.
- [138] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.
- [139] Yusuke T. Maeda and Masaki Sano. Regulatory dynamics of synthetic gene networks with positive feedback. *J. Mol. Biol.*, Vol. 359, No. 4, pp. 1107 – 1124, 2006.

- [140] Uri Alon. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.*, Vol. 8, pp. 450–461, 2007.
- [141] Shmoolik Mangan and Uri Alon. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. U.S.A.*, Vol. 100, No. 21, pp. 11980–11985, 2003.
- [142] Yves Meyer, Wafi Siala, Talaat Bashandy, Christophe Riondet, Florence Vignols, and Jean Philippe Reichheld. Glutaredoxins and thioredoxins in plants. *Biochim Biophys Acta Mol Cell Res*, Vol. 1783, No. 4, pp. 589–600, 2008.
- [143] Ruslan Deviatiiarov, Takahiro Kikawada, and Oleg Gusev. The complete mitochondrial genome of an anhydrobiotic midge *Polypedilum vanderplanki* (chironomidae, diptera). *Mitochondrial DNA A DNA Mapp Seq Anal*, Vol. 28, No. 2, pp. 218–220, 2017.
- [144] Roberto Mantovani. The molecular biology of the CCAAT-binding factor NF-Y. *Gene*, Vol. 239, No. 1, pp. 15–27, 1999.
- [145] Donald E Nelson, Peter P Repetti, Tom R Adams, Robert A Creelman, Jingrui Wu, David C Warner, Don C Anstrom, Robert J Bensen, Paolo P Castiglioni, Meghan G Donnarummo, et al. Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. *Proc. Natl. Acad. Sci. U.S.A.*, Vol. 104, No. 42, pp. 16450–16455, 2007.
- [146] Miao Chen, Yujuan Zhao, Chunliu Zhuo, Shaoyun Lu, and Zhenfei Guo. Overexpression of a NF-YC transcription factor from bermudagrass confers tolerance to drought and salinity in transgenic rice. *Plant Biotechnol. J.*, Vol. 13, No. 4, pp. 482–491, 2015.
- [147] Yoichiro Sogame, Jun Okada, Shingo Kikuta, Yugo Miyata, Richard Cornette, Oleg Gusev, and Takahiro Kikawada. Establishment of gene transfer and gene silencing methods in a desiccation-tolerant cell line, Pv11. *Extremophiles*, Vol. 21, No. 1, pp. 65–72, 2017.