

# 해수 온난화에 의한 열대 수산자원 영향 연구

Study of the Effect on Tropical Fishery Resources by Ocean Warming

유태우 응용사업팀

## 발간사

가속화되는 온난화 추세, 그에 따른 이상기후의 급증에 따라, 생물 피해가 급증하고 있다. 다수의 연구에 따르면 향후 엘니뇨/라니냐 등의 빈도가 증가할 것으로 예상되고, 이는 온도 변화에 취약한 생물들의 경작, 양식, 분포 등에 큰 악영향을 미칠 것으로 예상된다. 이런 경향은 UN의 sustainable development goal 17개 중 세 가지 목표(Zero hunger, Life below water, Life on land)로 제시될 정도로 점차 중요하게 인식되고 있고, 따라서 기후변화생물학에 대한 연구와 대책이 절실히 요구되고 있다.

엘니뇨에 따른 대규모 양식어패류의 폐사, 호주 Great Barrier Reef의 산호 및 어류들의 대규모 폐사 등이 매년 보고되고 있고, 해수의 평균온도 또한 차츰 증가하는 것으로 알려져 있다. 육상 생물에 비해 해양 생물은 적은 범위의 온도 변화가 이루어지는 환경에서 서식하고 진화되었기 때문에, 효율적으로 활동하고 번식할 수 있는 수온의 범위가 좁고, 또한 대부분이 변온동물로서 체온 조절 능력이 떨어지기 때문에, 온도 변화에 특히 취약하다. APEC 국가들, 특히 개발도상국들은 수산업 및 해양관광업에 국가경제 의존 비율이 높기 때문에, 수산 자원 보존을 위해, 수산자원의 적응 여부를 예측하는 연구가 절실히 요구된다. 해수 온도 상승이 생물들의 생존에 어떻게 영향을 미치는지, 적응 및 실패를 결정하는 생체 메커니즘이 어떻게 되는지에 대한 연구 결과가 차츰 축적되고 있다. 본 연구에서는 어류를 대상으로 하는 유전체 빅데이터를 수집하고 이 자료와 시스템생물학 방법을 이용해, 온난화 적응 여부를 판별할 수 있는 분석 파이프라인을 구축하였다. 이러한 시도 방법은 온난화 피해를 저감하기 위한 효율적인 형질을 가진 수산자원의 빠른 발굴

과 대책 마련, 경제 피해의 최소화와 같은 노력의 시발점이 될 것이다. 또한, 축적되는 기후변화생물학의 빅데이터 관리 및 활용 기술의 기준이 될 것이다.

연구 결과가 나오기 위해 본 연구를 수행한 유태우 박사의 노고 그리고 분석에 도움을 준 조원일 연구원에게 감사를 드린다. 또한 연구보고서의 품질을 제고하도록 훌륭한 논평을 해주신 익명의 검토자들께도 감사드린다. 본 연구보고서가 밑바탕이 되어 향후 관련된 연구들이 부족한 부분들을 꾸준히 채워나가길 바란다.

2018년 3월  
APEC 기후센터  
원장 정 홍 상

## ABSTRACT

다양한 온난화 시나리오들은 해양 온도의 꾸준한 상승 뿐 아니라 엘니뇨/라니냐 등의 이상기후에 따른 급격한 해양 온도 변화를 예견한다. 열대 해양의 변온생물들은 생명현상을 유지할 수 있는 한계온도에 가까운 환경에서 서식하며 체온 조절 능력이 약하기 때문에 기후 변화에 특히 취약하고, 이에 따라 다양성 및 개체 수 감소, 서식지 침입/이동 등이 예상된다. 해양 생물들은 수산업 및 관광업에 의존적인 열대/아열대 지역의 국가들에게 중요한 자원이기 때문에, 이에 대한 체계적인 연구가 필요하다. 각 생물집단의 기후 변화 적응 여부를 예측하기 위해서는 서식지 및 집단을 대상으로 온난화 시나리오에 따라 설계된 실험을 통해 적응력을 평가하여야 하지만, 이는 제어된 환경에서 수년간의 실험이 필요하기 때문에 현실적으로 무리가 있다. 이러한 단점을 극복하기 위한 방법으로 인간 질병 등의 연구에서 널리 쓰이는 유전자 마커를 이용한 온난화 적응 연구가 대체 수단으로 쓰일 수 있다. 본 연구에서는 수산자원의 빅데이터 분석을 통해 온도 변화에 취약성 여부를 예측하는 시스템을 구축하였다. High throughput sequencing 기술로 생성된 데이터를 수집/가공하였고, 단일 분석 파이프라인을 이용하여 차별발현유전자들을 찾아냈다. 또한 차별발현유전자들과 자동분류 알고리즘을 이용하여 기온 변화 적응 여부를 판별하는 분석 시스템을 만들었다. 빅데이터를 통해 수산자원의 온난화 적응 여부를 예측하는 본 연구는 수산/관광업에 의존하는 여러 국가들의 수산자원의 보존 연구를 위한 기반이 될 것이다.

# 목 차

1. 서론	1
2. 연구 자료 및 방법	3
2.1 데이터셋 수집	3
2.2 De novo assembly 및 유전자 분석	3
2.3 전사체 분석 및 기후변화 적응 분류 시스템 구현	4
3. 연구내용	5
3.1 데이터 수집	5
3.2 De novo assembly 및 유전자 예측	11
3.3 PCA 분석	12
3.4 DEG 분석	24
3.5 Classification System	39
4. 결론	43
▮ REFERENCES	44

## 표 목차

표 1. 수집된 데이터셋 요약 정보	6
표 2. 수집된 데이터셋의 Read와 Base의 통계	6
표 3. De Novo Assembly Statistics	12
표 4. Functional Annotation 결과	12
표 5. 데이터셋별 Treatment와 Adaptation 여부	15
표 6. 차별발현유전자 개수	25
표 7. 데이터셋 간의 공통인 차별발현유전자수	30
표 8. 모든 데이터셋에서 공통으로 나온 차별발현유전자와 기능 요약	30
표 9. Top 20 Biological Processes of Gene Ontology	39
표 10. SVM classification. TP = true positive; FN = false negative; FP = false positive; TN = true negative	40
표 11. Naive Bayes classification. TP = true positive; FN = false negative; FP = false positive; TN = true negative	41

---

## 그림 목차

그림 1. Normalization 전(왼쪽)과 후(오른쪽)의 read count 분포-----	13
그림 2. Principal component analysis. 검은 점은 온도변화에 잘 적응한 물고기 개체, 붉은 점은 적응하지 못한 개체를 나타낸다.-----	21
그림 3. Multidimensional scaling 분석. 검은 점은 온도변화에 잘 적응한 물고기 개체, 붉은 점은 적응하지 못한 개체를 나타낸다.-----	22
그림 4. 샘플간 distance matrix. 파란 색에 가까울수록 유사한 샘플, 빨간 색에 가까울수록 다른 샘플을 의미한다.-----	23
그림 5. 데이터셋별로 개별 DEG가 나타난 횟수-----	27
그림 6. SVM Classifier의 accuracy-----	40
그림 7. Naive Bayes classifier의 accuracy-----	41

## 1. 서론

수산자원은 어업, 관광업, 생태계 유지의 근간을 이룬다. 특히 2차, 3차 산업이 발달하지 못한 열대/아열대 지역의 개발도상국가들에서의 수산자원의 GDP 기여도는 더욱 중요하다. 수많은 기후변화 연구에 따르면 21세기 말까지 평균해양온도의 꾸준한 상승을 예견한다(Flato et al., 2013). 그러나 평균 온도 상승 뿐 아니라, 빈도증가가 예상되는 엘니뇨/라니냐에 따른 최저/최고 해수 온도의 급격한 변화로 인한 피해 또한 이미 급증하고 있다. 2016년 이상기후보고서(대한민국정부)에 따르면 엘니뇨에 따른 급격한 수온 증가로 양식분야에서만 500여억원의 피해가 발생하였다. 이는 해양생물이 대부분 변온생물로서 체온 조절 능력이 떨어지고, 해수의 온도 변화가 생체 기관에 그대로 영향을 주어, 유전물질 및 단백질의 기능에 악영향을 미치기 때문이다. 해수 온도가 각 생명체가 서식할 수 있는 한계온도 범위를 벗어날 경우, loss of equilibrium을 시작으로 죽음에 까지 이르게 된다. 이뿐만 아니라 한계온도 내에서도 최적의 서식 온도를 벗어날 경우, 발생, 대사, 생식 등에 영향을 미쳐 중장기적으로 성비 불균형, 개체수 감소, 집단의 몰락 등을 야기한다(Donelson et al., 2012; Lefevre, 2016).

수많은 학술 연구에서 해양생물들의 위기를 예측하고 있지만, 해양생물들이 기후변화에 적응을 할 수 있을지, 어느 정도로 적응을 할 수 있을지 연구는 현실적인 이유로 많이 되고 있지 못하다. 수산자원을 다년간 환경요인이 제어되는 실험조건에서 키우기 위해서는 비용, 공간, 숙련 인력이 필요하나, 예산 및 관심 부족으로 해당 연구가 이루어지지 못하는 것에 주로 기인한다. 이러한 한계를 극복하는 방법으로 유전물질 빅데이터를 이용한 종/집단별 기온변화 적응 능력을 판별하는 방법이 대두되고 있다. 이러한 방법론은 이미 인간 질병 예측에서 널리 쓰이기 시작한 것으로, 수천 달러 이내의 저비용과 수일~수주 내의 빠른 시간 안에 특정 질병에 걸릴 확률 등을 예측하여 미리 대비하게 하는 용도로 중요하게 사용되고 있다.

본 연구에서는 수산자원을 대상으로 기후변화 특히 온도 적응 여부를 판가름할 수 있는 분석 파이프라인을 구축하고자 하였다. 생체 데이터를 대규모로 생성해 내기 위한 다양한 종류의 High throughput sequencing 기술이 개발되어 왔지만, 최근에 널리 쓰여 범용성이 있고, 분석 파이프라인이 잘 개발되어 통용되는 Illumina의 HiSeq2000으로 생성된 어류의 기온 변화 적응 관련 데이터만을 대상으로 하였다. Genome

assembly를 이용하는 경우가 더 정확한 transcriptome(전사체) annotation과 정량화가 가능하지만, 비용이 많이 들고 시간이 오래 걸리는 단점이 있다. 반면 전사체만을 이용하는 경우는 중복 전사체 서열이 생성된다는 단점이 있지만, 비용과 시간이 크게 절감되기 때문에, 많은 생물들을 대상으로 하는 분석이 용이하다. 따라서, 본 연구에서는 전사체 분석 파이프라인을 구축하여 연구를 진행하였다. 다양한 온도 조건의 두 실험간 차별발현 유전자들을 찾아냈고, 이들을 분류 알고리즘의 feature로 사용하였다. 다차원 척도법을 통해 차별발현유전자를 이용하여 다른 온도 조건의 실험 샘플간에 분리가 잘 됨을 확인하였고, 기온 변화 적응 여부 판별 시스템에서도 쓰일 수 있음을 밝혔다.

생명체의 유전체 돌연변이 등을 통한 자연적인 적응은 진화의 시간 단위로 오래 걸리기 때문에, 적응력이 약하거나 강한 생물을 미리 판별하여 기후변화에 대비하는 연구와 정책이 필수적이다. 같은 종 내에서도 집단에 따라 환경에 대처하는 능력이 판이하게 다르기 때문에 기후 변화에 대한 적응 여부를 육안으로 구분/예측하는 것이 불가능하다. 따라서 생명 현상의 근원인 유전자 레벨에서 이를 판별하는 과정이 필요하다. 빅데이터를 이용한 체계화된 분석 방법을 제공한 본 연구는 수산자원 뿐 아니라 생태계의 종 보존 연구를 위한 기반이 될 것이다.

## 2. 연구 자료 및 방법

### 2.1 데이터셋 수집

본 연구에서 사용된 데이터셋의 종류는 다음과 같다: 1) catfish(accession number: SRA049427); 2) zebrafish(accession number: E-MTAB-1155); 3) salmon(accession number: GSE59756); 4) tilapia and zebrafish(accession number: GSE69965); 5) damselfish(accession number: PRJNA255544). 각 데이터셋 별로 다양한 실험 조건 하의 여러 샘플이 존재하고, 각 샘플별로 Fastq 형식의 read 들을 NCBI와 EBI의 데이터 저장소에서 다운로드하였다. 이 외에도 몇 가지 데이터셋을 추가로 찾아냈지만, 샘플 수의 부족 및 데이터 보관이 잘 되어 있지 않아 제외하였다. 수집된 데이터셋은 in-house 데이터베이스화하였다.

### 2.2 *De novo* assembly 및 유전자 분석

Zebrafish나 damselfish 등 몇 종에 대해서는 이미 고품질의 게놈 서열이 존재한다. 그러나, 본 연구의 목표는 전사체만을 이용한 바이오마커 발굴과 온난화적응 판별시스템 구축이기 때문에, 각 종별로 전사체만을 이용하여 assembly를 만들었다.

RNA-Seq 기술로 생성된 read 데이터에는 adapter 및 sequencing error 등이 존재한다. 이로 인한 noisy base를 제거하고 기준 품질 이상의 read들만을 사용하기 위해 trimmomatic v0.36(Bolger et al., 2014)이 기본 옵션으로 적용되었다. 이 결과 생성된 수백~수천만개의 짧은 read들을 조립하여 mRNA서열을 만드는 과정(*de novo* assembly)를 위해 trinity v2.3.2(Grabherr et al., 2011)가 사용되었다.

Assembly이후에는 수십만개의 contig들이 생성되고 이 중 많은 부분은 assembly 중간에 생성되는 artifact이거나, 같은 transcript로부터 나오는 중복서열이다. 이들을 제거하기 위해, Transdecoder v4.1(Haas et al., 2013)을 이용하여 assembly 결과로 생성된 contig에서 coding region 후보를 추출하였고, 중복 서열은 cdhit-est v4.6.7(Li and Godzik, 2006)을 이용하여 제거되었다(95% similarity threshold).

최종 transcript의 유전자 이름 지정 및 기능 유추를 위해 blastp v2.6.0(Altschul et al., 1990)과 UniProtKB/Swiss-Prot(Boutet et al., 2007)를 사용하였다.

## 2.3 전사체 분석 및 기후변화 적응 분류 시스템 구현

유전자별 mRNA 발현량을 측정하기 위하여, 전처리된 read들을 각 유전자 서열에 Bowtie2 v2.2.9(Langmead and Salzberg, 2012)를 이용하여 mapping하였고, 이 결과 생성되는 SAM파일과 SamTools v1.4.1(Li et al., 2009)을 이용하여 발현량을 정량화하였다. 각 데이터셋별로 DESeq2 v1.16.1(Love et al., 2014)를 이용해 normalize 및 differentially expressed genes(DEG)를 찾아냈다. Principal component analysis(PCA)와 multidimensional scaling(MDS) 분석은 R v3.4.0을 이용하였다.

온도 변화에 어류가 적응하는 여부를 전사체를 이용하여 판별해주는 분류 시스템 구축을 위해, 각 데이터셋의 샘플별로 온도 적응 여부를 Yes 혹은 No로 정의하고, R의 naive Bayes, support vector machine(svm), random forest 기법을 이용하여 training과 test를 진행하였다.

### 3. 연구내용

#### 3.1 데이터 수집

수산자원 온난화와 관련된 문헌 검색 결과, 대부분 연구에서는 Microarray 혹은 Illumina HiSeq2000 등의 기술로 전사체(Transcriptome)를 시퀀싱하였다. 그러나 최근 연구에서는 범용성이 좋고 데이터 생성량이 더 많은 HiSeq2000이 더 많이 사용되었다. 따라서 본 연구에서는 HiSeq2000 기술로 생성된 데이터만을 사용하였다. 메기 (catfish), 연어(salmon), 역돔(tilapia) 등의 식용물고기를 비롯해, 분자생물학/생태학 연구가 많이 되어있는 제브라피시(zebrafish)와 열대자리돔(damselfish)의 다양한 조직, 발생단계, 온도에 대한 데이터가 수집되었다(표 1). 메기 데이터는 24°C의 온도에서 적응된 성체를 36°C에서 heat stress를 주어 평형상태를 잃은 개체들을 수집하여 간과 아가미로부터 전사체를 추출하여 24°C에서만 적응된 대조군과 비교한 것이다. 제브라피시(E-MTAB-1155 데이터)의 경우, 배아 상태에서 22, 27, 32°C 세 온도에서 길러진 개체들을 성체가 된 후 16, 22, 27, 32, 34°C의 여러 온도로 나누어서 하루간 길러본 후에 생리반응과 근육에서의 전사체 반응을 살펴본 것이다. 연어의 경우, 12°C에서 길러진 개체들을 12, 15, 18, 21, 25°C 탱크로 나누어 옮겨 3시간 동안 적응시킨 후, 12°C 탱크로 다시 옮겨 아가미에서의 전사체 반응을 살펴보았다. GSE69965 데이터셋은 제브라피시와 대표적 민물양식어종인 나일틸라피아를 대상으로 28°C에서 5세대 이상 길러진 개체들을 8°C의 저온에서 기른 후 아가미의 전사체 반응을 살펴보고 각 종에서의 저온 적응 반응 및 종간의 비교를 연구한 경우이다. 열대자리돔 자료는 온난화 적응 연구가 잘 이루어진 어류로서, 다양한 온도 조건 및 위도가 다른 두 지역으로부터 두 세대에 걸쳐 많은 샘플(64개체)가 얻어져, 다양한 비교 분석이 가능한 데이터셋이다.

표 1. 수집된 데이터셋 요약 정보

일반명	Accession Number	# of Samples	Tissue	참고문헌
catfish	SRA049427	6	liver and gill	Liu et al., 2013
zebrafish	E-MTAB-1155	22	muscle	Scott and Johnson, 2012
salmon	GSE59756	15	gill	Tomalty et al., 2015
tilapia	GSE69965	12	gill	Hu et al., 2016
zebrafish	GSE69965	12	gill	Hu et al., 2016
damselish	PRJNA255544	64	liver	Veilleux et al., 2015

표 2. 수집된 데이터셋의 Read와 Base의 통계

데이터셋	샘플	Read 개수	Base 개수
SRA049427 (catfish)	SRR917955	56,639,172	5,720,556,372
	SRR917956	72,870,754	7,359,946,154
	SRR917957	51,869,048	5,238,773,848
	SRR917958	46,578,524	4,704,430,924
	SRR917959	66,622,238	6,728,846,038
	SRR917960	63,810,776	6,444,888,376
E-MTAB-1155 (zebrafish)	ERR145631	76,578,926	7,734,471,526
	ERR145632	59,488,960	6,008,384,960
	ERR145633	60,855,640	6,146,419,640
	ERR145634	55,914,704	5,647,385,104
	ERR145635	61,413,038	6,202,716,838
	ERR145636	76,712,998	7,748,012,798
	ERR145637	20,834,580	2,104,292,580
	ERR145638	64,404,156	6,504,819,756
	ERR145639	72,498,924	7,322,391,324
	ERR145646	8,048,348	812,883,148
	ERR145647	106,615,294	10,768,144,694
	ERR145648	3,276,774	330,954,174
	ERR145649	75,597,756	7,635,373,356
	ERR145650	2,772,654	280,038,054
	ERR145651	46,920,472	4,738,967,672
	ERR145652	2,275,332	229,808,532
ERR145653	43,046,486	4,347,695,086	
ERR145654	801,470	80,948,470	

표 2. 수집된 데이터셋의 Read와 Base의 통계(계속)

데이터셋	샘플	Read 개수	Base 개수
E-MTAB-1155 (zebrafish)	ERR145655	63,920,168	6,455,936,968
	ERR147026	29,757,664	3,005,524,064
	ERR147027	34,217,722	3,455,989,922
	ERR147028	31,654,494	3,197,103,894
GSE59756 (salmon)	SRR1524980	122,203,854	12,342,589,254
	SRR1524981	139,832,444	14,123,076,844
	SRR1524982	123,082,022	12,431,284,222
	SRR1524983	137,187,906	13,855,978,506
	SRR1524984	97,658,404	9,863,498,804
	SRR1524985	91,191,358	9,210,327,158
	SRR1524986	119,896,152	12,109,511,352
	SRR1524987	137,220,530	13,859,273,530
	SRR1524988	121,966,874	12,318,654,274
	SRR1524989	101,028,102	10,203,838,302
	SRR1524990	131,782,048	13,309,986,848
	SRR1524991	124,193,658	12,543,559,458
	SRR1524992	106,720,008	10,778,720,808
	SRR1524993	109,707,970	11,080,504,970
SRR1524994	106,896,038	10,796,499,838	
GSE69965 (zebrafish)	SRR2067881	26,310,104	2,657,320,504
	SRR2067882	24,046,954	2,428,742,354
	SRR2067883	29,854,060	3,015,260,060
	SRR2067884	30,874,104	3,118,284,504
	SRR2067886	30,405,560	3,070,961,560
	SRR2067887	25,773,114	2,603,084,514
	SRR2067888	31,503,608	3,181,864,408
	SRR3579895	24,385,010	2,462,886,010
	SRR3579896	29,426,658	2,972,092,458
	SRR3579897	29,384,866	2,967,871,466
	SRR3579898	30,244,142	3,054,658,342
	SRR3579899	33,586,050	3,392,191,050
GSE69965 (tilapia)	SRR2067864	23,882,388	2,412,121,188
	SRR2067865	28,263,450	2,854,608,450
	SRR2067866	28,590,338	2,887,624,138

표 2. 수집된 데이터셋의 Read와 Base의 통계(계속)

데이터셋	샘플	Read 개수	Base 개수
GSE69965 (tilapia)	SRR2067867	31,351,944	3,166,546,344
	SRR2067868	26,855,700	2,712,425,700
	SRR2067869	34,053,640	3,439,417,640
	SRR2067870	29,712,256	3,000,937,856
	SRR2067873	30,430,518	3,073,482,318
	SRR2067874	24,852,508	2,510,103,308
	SRR3579892	18,057,772	1,823,834,972
	SRR3579893	22,125,982	2,234,724,182
	SRR3579894	26,495,110	2,676,006,110
PRJNA255544 (damselfish)	L1_HA1	29,749,092	3,004,658,292
	L1_HA2	23,750,660	2,398,816,660
	L1_HAB1	23,020,828	2,325,103,628
	L1_HAB2	3,607,554	364,362,954
	L1_HB1	34,216,192	3,455,835,392
	L1_HB2	71,301,422	7,201,443,622
	L1_HC1	81,051,812	8,186,233,012
	L1_HC2	22,327,056	2,255,032,656
	L2_HAC1	25,808,902	2,606,699,102
	L2_HAC2	18,069,500	1,825,019,500
	L2_OA1	21,907,454	2,212,652,854
	L2_OA4	25,085,458	2,533,631,258
	L2_OB2	25,950,338	2,620,984,138
	L2_OB3	21,973,920	2,219,365,920
	L2_OC4	31,634,230	3,195,057,230
	L2_OC5	38,237,840	3,862,021,840
	L3_OAB1	63,280,392	6,391,319,592
	L3_OAB2	26,107,126	2,636,819,726
	L3_OABY1	27,221,870	2,749,408,870
	L3_OABY2	28,118,052	2,839,923,252
	L3_OAC1	67,929,814	6,860,911,214
	L3_OAC2	77,268,814	7,804,150,214
	L3_OACY1	33,936,470	3,427,583,470
	L3_OACY2	24,610,518	2,485,662,318
L5_A1	22,063,194	2,228,382,594	

표 2. 수집된 데이터셋의 Read와 Base의 통계(계속)

데이터셋	샘플	Read 개수	Base 개수
PRJNA255544 (damselfish)	L5_A2	58,489,494	5,907,438,894
	L5_B1	32,918,234	3,324,741,634
	L5_B2	29,446,504	2,974,096,904
	L5_C1	62,174,698	6,279,644,498
	L5_C2	23,761,356	2,399,896,956
	L5_tB1	26,279,826	2,654,262,426
	L5_tB2	25,174,968	2,542,671,768
	L6_LAB1	35,715,454	3,607,260,854
	L6_LAB2	19,609,792	1,980,588,992
	L6_LAC1	21,561,986	2,177,760,586
	L6_LAC2	16,591,848	1,675,776,648
	L6_sC1	8,527,244	861,251,644
	L6_sC2	25,767,870	2,602,554,870
	L6_tC1	32,590,698	3,291,660,498
	L6_tC2	22,117,190	2,233,836,190
	L7_aB1	25,129,796	2,538,109,396
	L7_aB2	17,109,228	1,728,032,028
	L7_LtBA1	33,179,332	3,351,112,532
	L7_LtBC1	52,085,434	5,260,628,834
	L7_LtBC2	42,035,586	4,245,594,186
	L7_LtCA1	56,626,328	5,719,259,128
	L7_LtCA2	25,705,186	2,596,223,786
	L8_aAC1	40,332,924	4,073,625,324
	L8_aAC2	30,237,698	3,054,007,498
	L8_atBC1	56,665,560	5,723,221,560
	L8_atBC2	30,948,916	3,125,840,516
	L8_atCA1	37,051,912	3,742,243,112
	L8_atCA2	32,781,558	3,310,937,358
	L8_HA1F	39,558,698	3,995,428,498
	L8_HA2F	23,690,064	2,392,696,464
	HAB2F_2834	18,932,308	1,912,163,108
	HAB3F_2835	11,794,514	1,191,245,914
	HAC1F_2836	9,192,832	928,476,032
	HAC2F_2837	10,714,304	1,082,144,704

표 2. 수집된 데이터셋의 Read와 Base의 통계(계속)

데이터셋	샘플	Read 개수	Base 개수
PRJNA255544 (damselfish)	HB1F_2830	13,251,772	1,338,428,972
	HB2F_2831	10,572,402	1,067,812,602
	HC1F_2833	18,393,714	1,857,765,114
	HC2F_2832	19,622,050	1,981,827,050
	LtBA3_2815a	162,557,386	16,418,295,986
	A1	62,935,824	6,356,518,224
	A3	53,275,942	5,380,870,142
	A4	54,812,310	5,536,043,310
	A5	59,542,406	6,013,783,006
	B3	64,601,812	6,524,783,012
	B4	62,992,254	6,362,217,654
	B5	60,330,574	6,093,387,974
	C3	65,196,926	6,584,889,526
	C4	63,334,944	6,396,829,344
	C5	63,319,688	6,395,288,488
	C6	47,482,514	4,795,733,914
	HA1	66,858,670	6,752,725,670
	HA3F	60,825,280	6,143,353,280
	HA3	66,386,598	6,705,046,398
	HA4	61,148,896	6,176,038,496
	HAB1F	68,689,036	6,937,592,636
	HAC3	76,881,792	7,765,060,992
	HAC4	72,516,090	7,324,125,090
	HB3	73,894,888	7,463,383,688
	HC3F	75,039,380	7,578,977,380
	HC3	83,140,044	8,397,144,444
	HC4	79,753,934	8,055,147,334
	OA1F	65,270,940	6,592,364,940
	OA1	64,837,734	6,548,611,134
	OA3	66,913,670	6,758,280,670
	OA4	62,776,166	6,340,392,766
	OA5	51,238,844	5,175,123,244
	OAB1F	57,702,200	5,827,922,200
OAB3	62,357,496	6,298,107,096	

표 2. 수집된 데이터셋의 Read와 Base의 통계(계속)

데이터셋	샘플	Read 개수	Base 개수
PRJNA255544 (damselselfish)	OAB4	64,924,890	6,557,413,890
	OAC1F	60,715,830	6,132,298,830
	OAC3	63,716,342	6,435,350,542
	OAC4	64,583,798	6,522,963,598
	OB1	56,699,314	5,726,630,714
	OB4	60,380,002	6,098,380,202
	OB5	55,314,988	5,586,813,788
	OC1F	60,103,502	6,070,453,702
	OC1	63,504,216	6,413,925,816
	OC2	65,583,984	6,623,982,384
	OC3	56,088,668	5,664,955,468
	sC3	49,958,226	5,045,780,826
	sC4	54,412,158	5,495,627,958
	tB3	67,196,138	6,786,809,938
	tB4	57,102,790	5,767,381,790
	tC3	54,125,144	5,466,639,544
	tC4	58,794,100	5,938,204,100
	tC5	64,107,530	6,474,860,530

### 3.2 *De novo* assembly 및 유전자 예측

본 연구의 목적은 RNA-Seq을 이용한 온난화 적응 여부를 판별하는 분류 시스템 구축이 목적이기 때문에, genome assembly가 되어 있는 종의 경우에도 *de novo* assembly로 만들어진 transcriptome 분석에만 집중하였다. Assembly로 생성된 수많은 contig들로부터 transdecoder와 cdhit-est를 이용하여 coding region 후보만 추출하는 filtering 과정을 거쳤다. 각 데이터셋의 assembly 통계 및 최종 contig 갯수는 다음과 같다.

표 3. *De Novo* Assembly Statistics

데이터셋	# of Contigs	N50	Maximum Length	# of Final Contigs
SRA049427 (catfish)	215,515	1,659	23,139	29,486
E-MTAB-1155 (zebrafish)	255,685	1,179	23,097	31,273
GSE59756 (salmon)	809,449	1,338	18,679	71,250
GSE69965 (tilapia)	421,330	2,615	24,196	59,041
GSE69965 (zebrafish)	494,584	1,199	38,195	46,867
PRJNA255544 (damsel fish)	474,964	872	16,059	41,176

최종 contig set은 해당 실험으로부터 나온 일정 발현량 이상의 유전자 서열로 간주할 수 있으며, A/T/G/C 4개 염기의 1차원적 나열 형태이기 때문에, blast 등을 이용한 homology search 등을 통한 functional annotation 과정이 필요하다. 본 연구에서는 널리 쓰이는 NCBI non-redundant data보다 적은 크기의 데이터베이스이긴 하지만, 유전자의 기능이 잘 annotation되어 있는 UniProt-KB를 이용하여 contig들의 기능을 유추하였다. 표 4에 나타난 것처럼 76~87%의 contig들이 최소 1개 이상의 UniProt entry에 유의미한 유사성을 보였고, 이 비율은 전사체의 assembly가 잘 이루어졌음을 알려준다.

표 4. Functional Annotation 결과

데이터셋	# of Annotated Transcript	# of Hit in UniProt
SRA049427 (catfish)	25,681	256,549
E-MTAB-1155 (zebrafish)	26,512	252,403
GSE59756 (salmon)	59,177	262,339
GSE69965 (zebrafish)	37,832	263,460
GSE69965 (tilapia)	45,225	268,068
PRJNA255544 (damsel fish)	31,374	239,541

### 3.3 PCA 분석

High throughput 데이터를 이용한 연구에서 실험 조건간의 차이점을 보는 방법으로 Principal component analysis나 Multidimensional scaling과 같은 차원변환을 이

용한 기법이 많이 사용된다. 전사체의 발현량을 feature로 사용하여 샘플간에 분리가 일어나는지를 확인하였다. 우선, 다른 데이터셋 및 샘플간 비교의 공정함을 위해 데이터셋 별로 Transcripts Per Kilobase Million(TPM) normalization과정을 거쳤다(그림 1). 각 샘플별 온도 변화 적응 유무는 각 데이터셋의 해당 논문으로부터 얻어졌다(표 5).

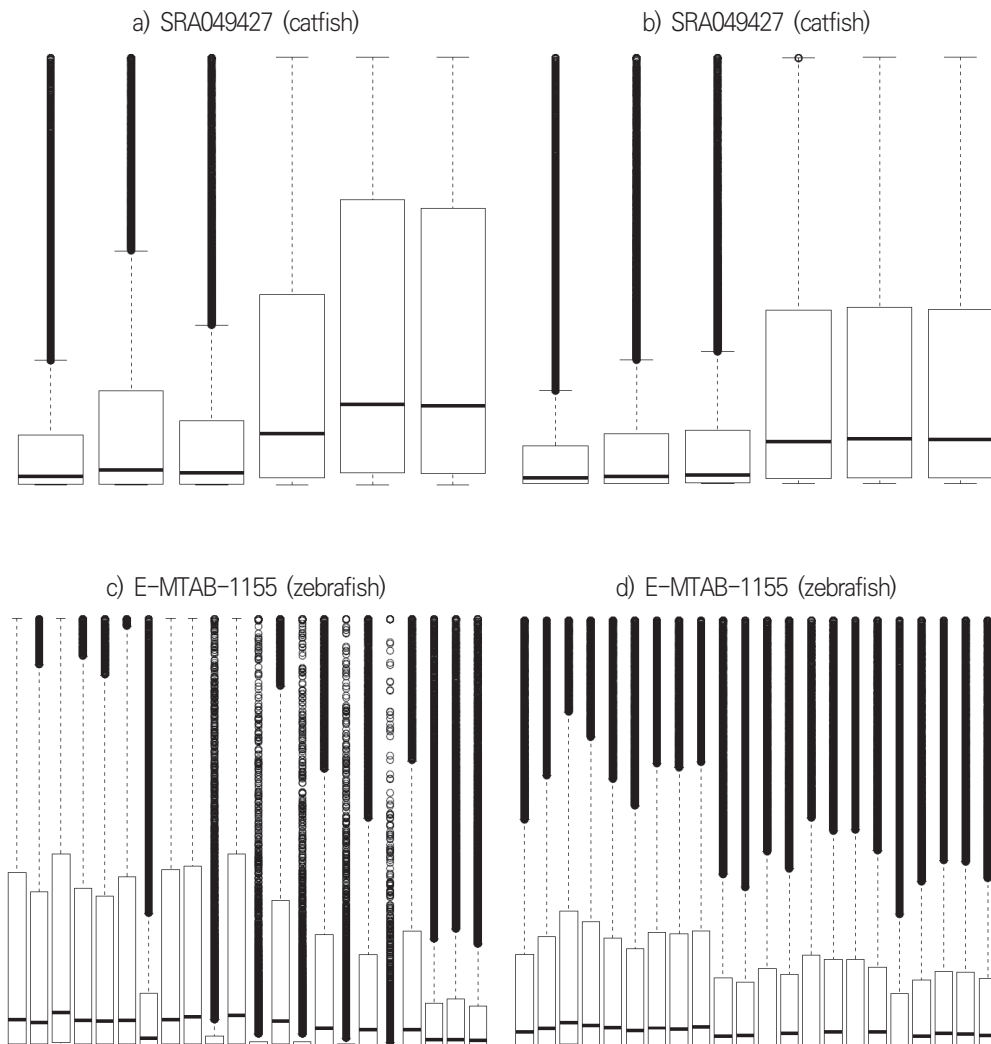


그림 1. Normalization 전(왼쪽)과 후(오른쪽)의 read count 분포

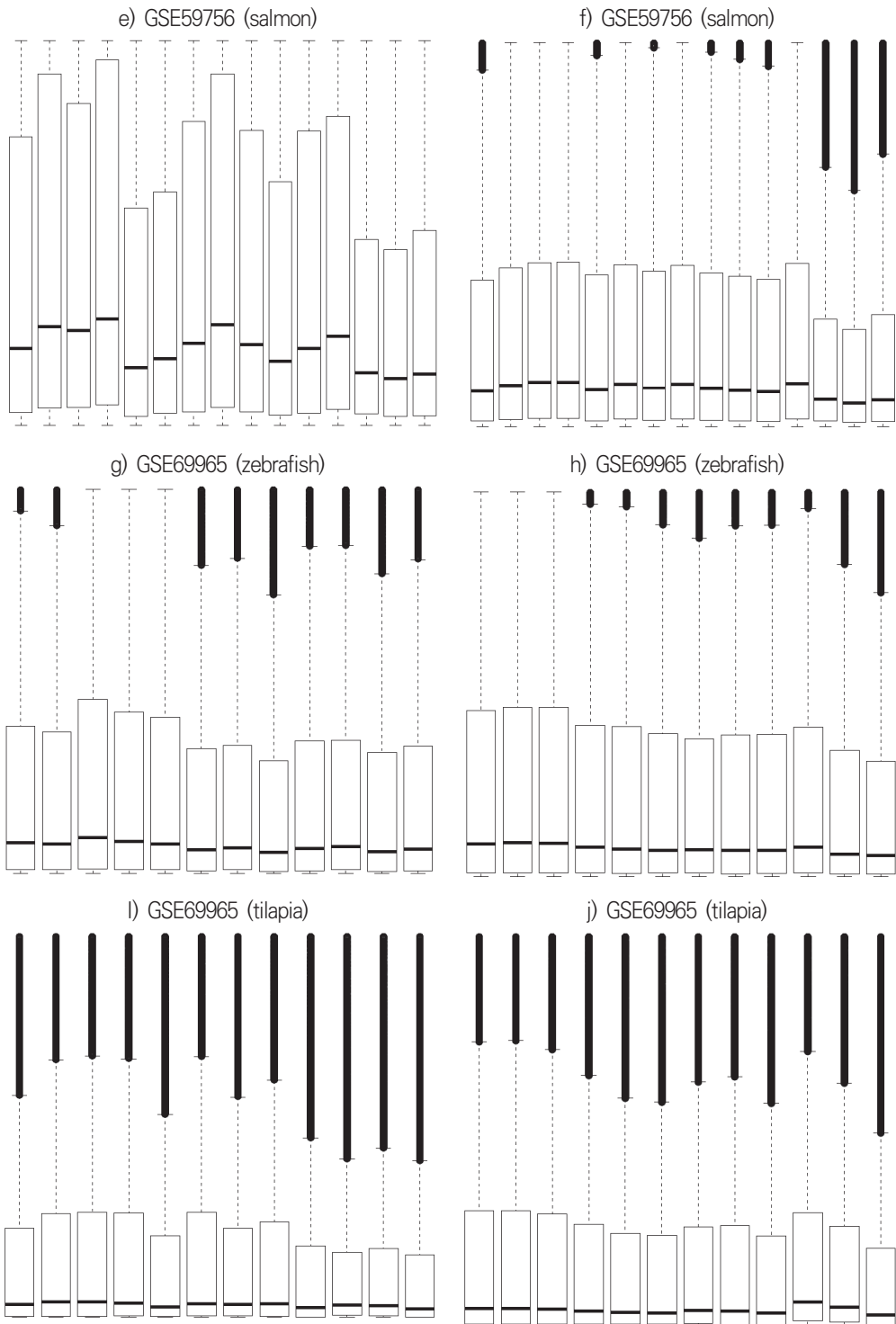


그림 1. Normalization 전(왼쪽)과 후(오른쪽)의 read count 분포(계속)

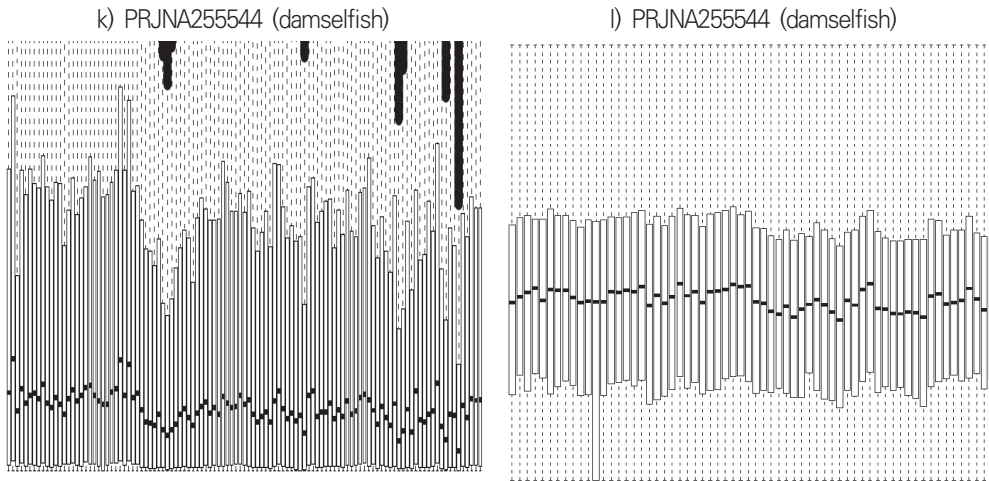


그림 1. Normalization 전(왼쪽)과 후(오른쪽)의 read count 분포(계속)

표 5. 데이터셋별 Treatment와 Adaptation 여부

데이터셋	ID	Treatment	Adaptation
SRA049427	SRR917955	liver_control	○
	SRR917956	liver_intolerant	×
	SRR917957	liver_tolerant	○
	SRR917958	gill_control	○
	SRR917959	gill_intolerant	×
	SRR917960	gill_tolerant	○
GSE59756	SRR1524994	gill_25C	×
	SRR1524993	gill_25C	×
	SRR1524992	gill_25C	×
	SRR1524991	gill_21C	×
	SRR1524990	gill_21C	×
	SRR1524989	gill_21C	×
	SRR1524988	gill_18C	○
	SRR1524987	gill_18C	○
	SRR1524986	gill_18C	○
	SRR1524985	gill_15C	○
	SRR1524984	gill_15C	○
	SRR1524983	gill_15C	○
	SRR1524982	gill_12C	○

표 5. 데이터셋별 Treatment와 Adaptation 여부(계속)

데이터셋	ID	Treatment	Adaptation
GSE59756	SRR1524981	gill_12C	○
	SRR1524980	gill_12C	○
GSE69965_Danio_rerio	SRR2067881	gill_28C	○
	SRR2067882	gill_28C	○
	SRR2067883	gill_28C	○
	SRR2067884	gill_8C_0h_LOE	×
	SRR2067886	gill_8C_0h_LOE	×
	SRR2067887	gill_8C_12h_LOE	×
	SRR2067888	gill_8C_12h_LOE	×
	SRR3579895	gill_8C_0h	○
	SRR3579896	gill_8C_6h	○
	SRR3579897	gill_8C_6h	○
	SRR3579898	gill_8C_6h	○
	SRR3579899	gill_8C_12h	○
	GSE69965_Oreochromis_niloticus	SRR3579894	gill_8C_12h
SRR3579893		gill_8C_6h	○
SRR3579892		gill_8C_0h	○
SRR2067874		gill_8C_12h_LOE	×
SRR2067873		gill_8C_12h_LOE	×
SRR2067870		gill_8C_6h_LOE	×
SRR2067869		gill_8C_6h_LOE	×
SRR2067868		gill_8C_0h_LOE	×
SRR2067867		gill_8C_0h_LOE	×
SRR2067866		gill_28C	○
SRR2067865		gill_28C	○
SRR2067864		gill_28C	○
E-MTAB-1155		ERR147028	muscle_E32C_A27C
	ERR147027	muscle_E32C_A27C	○
	ERR147026	muscle_E32C_A27C	○
	ERR145655	muscle_E27C_A27C	×
	ERR145654	muscle_E27C_A27C	×
	ERR145653	muscle_E27C_A27C	×
	ERR145652	muscle_E27C_A27C	×

표 5. 데이터셋별 Treatment와 Adaptation 여부(계속)

데이터셋	ID	Treatment	Adaptation
E-MTAB-1155	ERR145651	muscle_E27C_A27C	×
	ERR145650	muscle_E27C_A27C	×
	ERR145649	muscle_E27C_A27C	×
	ERR145648	muscle_E27C_A27C	×
	ERR145647	muscle_E32C_A27C	○
	ERR145646	muscle_E32C_A27C	○
	ERR145639	muscle_E27C_A16C	×
	ERR145638	muscle_E27C_A16C	×
	ERR145637	muscle_E27C_A16C	×
	ERR145636	muscle_E27C_A16C	×
	ERR145635	muscle_E27C_A16C	×
	ERR145634	muscle_E32C_A16C	○
	ERR145633	muscle_E32C_A16C	○
	ERR145632	muscle_E32C_A16C	○
	ERR145631	muscle_E32C_A16C	○
	PRJNA255544_damselfish	HAB2F_2834	heron_F1_liver_acute_0_1.5C
HAB3F_2835		heron_F1_liver_acute_0_1.5C	○
HAC1F_2836		heron_F1_liver_acute_0_3C	○
HAC2F_2837		heron_F1_liver_acute_0_3C	○
HAC3		heron_F1_liver_acute_0_3C	○
HAC4		heron_F1_liver_acute_0_3C	○
HB1F_2830		heron_F1_liver_dev_1.5C	○
HB2F_2831		heron_F1_liver_dev_1.5C	○
HC1F_2833		heron_F1_liver_dev_3C	○
HC2F_2832		heron_F1_liver_dev_3C	○
HA1		heron_F1_liver_control	○
L1_HA1		heron_F1_liver_control	○
L1_HA2		heron_F1_liver_control	○
L8_HA1F		heron_F1_liver_control	○
L8_HA2F		heron_F1_liver_control	○
HA3		heron_F1_liver_control	○
HA3F		heron_F1_liver_control	○
HA4		heron_F1_liver_control	○

표 5. 데이터셋별 Treatment와 Adaptation 여부(계속)

데이터셋	ID	Treatment	Adaptation
PRJNA255544_damselfish	HAB1F	heron_F1_liver_acute_0_1.5C	○
	L1_HAB1	heron_F1_liver_acute_0_1.5C	○
	L1_HAB2	heron_F1_liver_acute_0_1.5C	○
	L1_HB1	heron_F1_liver_dev_1.5C	○
	L1_HB2	heron_F1_liver_dev_1.5C	○
	HB3	heron_F1_liver_dev_1.5C	○
	L1_HC1	heron_F1_liver_dev_3C	○
	L1_HC2	heron_F1_liver_dev_3C	○
	HC3	heron_F1_liver_dev_3C	○
	HC3F	heron_F1_liver_dev_3C	○
	HC4	heron_F1_liver_dev_3C	○
	L2_HAC1	heron_F1_liver_acute_0_3C	○
	L2_HAC2	heron_F1_liver_acute_0_3C	○
	OA1	palm_F1_liver_control	○
	L2_OA1	palm_F1_liver_control	○
	OA1F	palm_F1_liver_control	○
	OA4	palm_F1_liver_control	○
	L2_OA4	palm_F1_liver_control	○
	OA3	palm_F1_liver_control	○
	OA5	palm_F1_liver_control	○
	OB2	palm_F1_liver_dev_1.5C	×
	OB3	palm_F1_liver_dev_1.5C	×
	L2_OB2	palm_F1_liver_dev_1.5C	×
	L2_OB3	palm_F1_liver_dev_1.5C	×
	OB1	palm_F1_liver_dev_1.5C	×
	OB4	palm_F1_liver_dev_1.5C	×
	OB5	palm_F1_liver_dev_1.5C	×
	OC1	palm_F1_liver_dev_3C	×
	OC1F	palm_F1_liver_dev_3C	×
	OC2	palm_F1_liver_dev_3C	×
	OC3	palm_F1_liver_dev_3C	×
	OC4	palm_F1_liver_dev_3C	×
	OC5	palm_F1_liver_dev_3C	×

표 5. 데이터셋별 Treatment와 Adaptation 여부(계속)

데이터셋	ID	Treatment	Adaptation
PRJNA255544_damselfish	L2_OC4	palm_F1_liver_dev_3C	×
	L2_OC5	palm_F1_liver_dev_3C	×
	A1	palm_F2_liver_control	○
	L5_A1	palm_F2_liver_control	○
	L5_A2	palm_F2_liver_control	○
	A3	palm_F2_liver_control	○
	A4	palm_F2_liver_control	○
	A5	palm_F2_liver_control	○
	B1	palm_F2_liver_dev_1.5C	×
	L5_B1	palm_F2_liver_dev_1.5C	×
	L5_B2	palm_F2_liver_dev_1.5C	×
	B2	palm_F2_liver_dev_1.5C	×
	B3	palm_F2_liver_dev_1.5C	×
	B4	palm_F2_liver_dev_1.5C	×
	B5	palm_F2_liver_dev_1.5C	×
	C1	palm_F2_liver_dev_3C	×
	C2	palm_F2_liver_dev_3C	×
	C3	palm_F2_liver_dev_3C	×
	C4	palm_F2_liver_dev_3C	×
	C5	palm_F2_liver_dev_3C	×
	C6	palm_F2_liver_dev_3C	×
	L5_C1	palm_F2_liver_dev_3C	×
	L5_C2	palm_F2_liver_dev_3C	×
	L5_tB1	palm_F2_liver_trans_1.5C	○
	L5_tB2	palm_F2_liver_trans_1.5C	○
	tB1	palm_F2_liver_trans_1.5C	○
	tB2	palm_F2_liver_trans_1.5C	○
	tB3	palm_F2_liver_trans_1.5C	○
	tB4	palm_F2_liver_trans_1.5C	○
	sC1	palm_F2_liver_step_3C	○
	sC2	palm_F2_liver_step_3C	○
	sC3	palm_F2_liver_step_3C	○
	sC4	palm_F2_liver_step_3C	○

표 5. 데이터셋별 Treatment와 Adaptation 여부(계속)

데이터셋	ID	Treatment	Adaptation
PRJNA255544_damselfish	L6_sC1	palm_F2_liver_step_3C	○
	L6_sC2	palm_F2_liver_step_3C	○
	tC1	palm_F2_liver_trans_3C	○
	L6_tC1	palm_F2_liver_trans_3C	○
	L6_tC2	palm_F2_liver_trans_3C	○
	tC2	palm_F2_liver_trans_3C	○
	tC3	palm_F2_liver_trans_3C	○
	tC4	palm_F2_liver_trans_3C	○
	tC5	palm_F2_liver_trans_3C	○
	HA1F	heron_F1_liver_control	○
	HA2F	heron_F1_liver_control	○

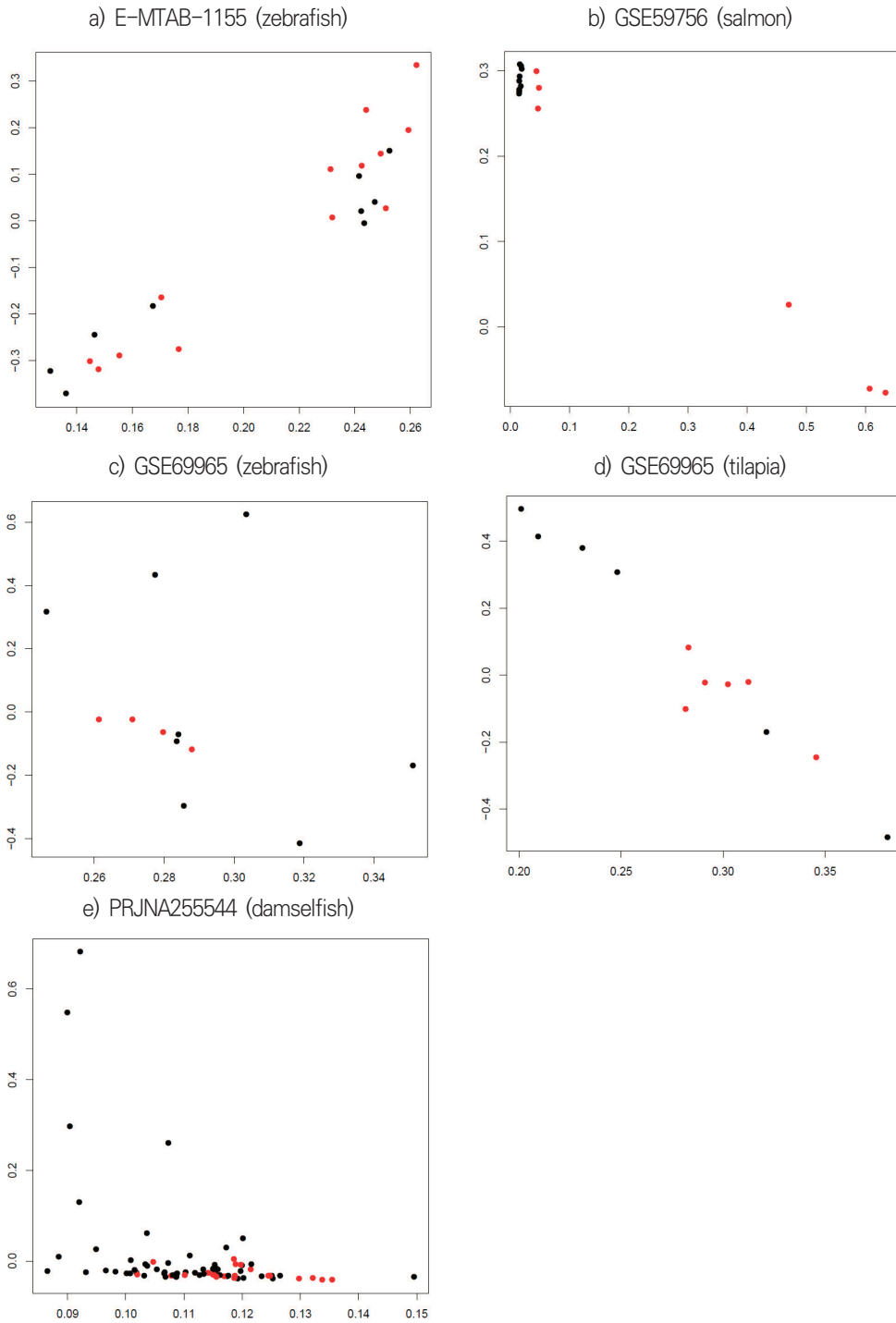


그림 2. Principal component analysis. 검은 점은 온도변화에 잘 적응한 물고기 개체, 붉은 점은 적응하지 못한 개체를 나타낸다.

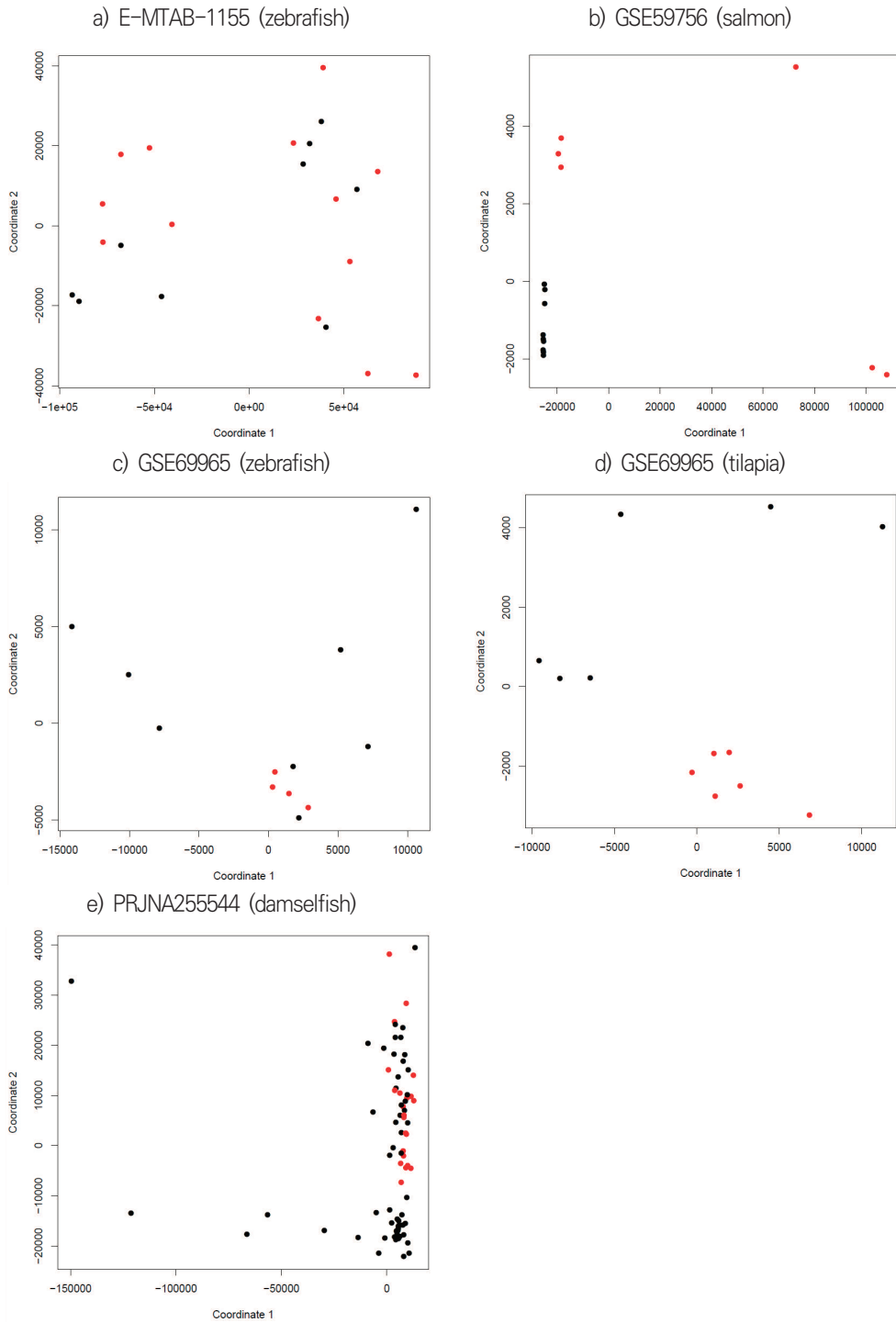


그림 3. Multidimensional scaling 분석. 검은 점은 온도변화에 잘 적응한 물고기 개체, 붉은 점은 적응하지 못한 개체를 나타낸다.

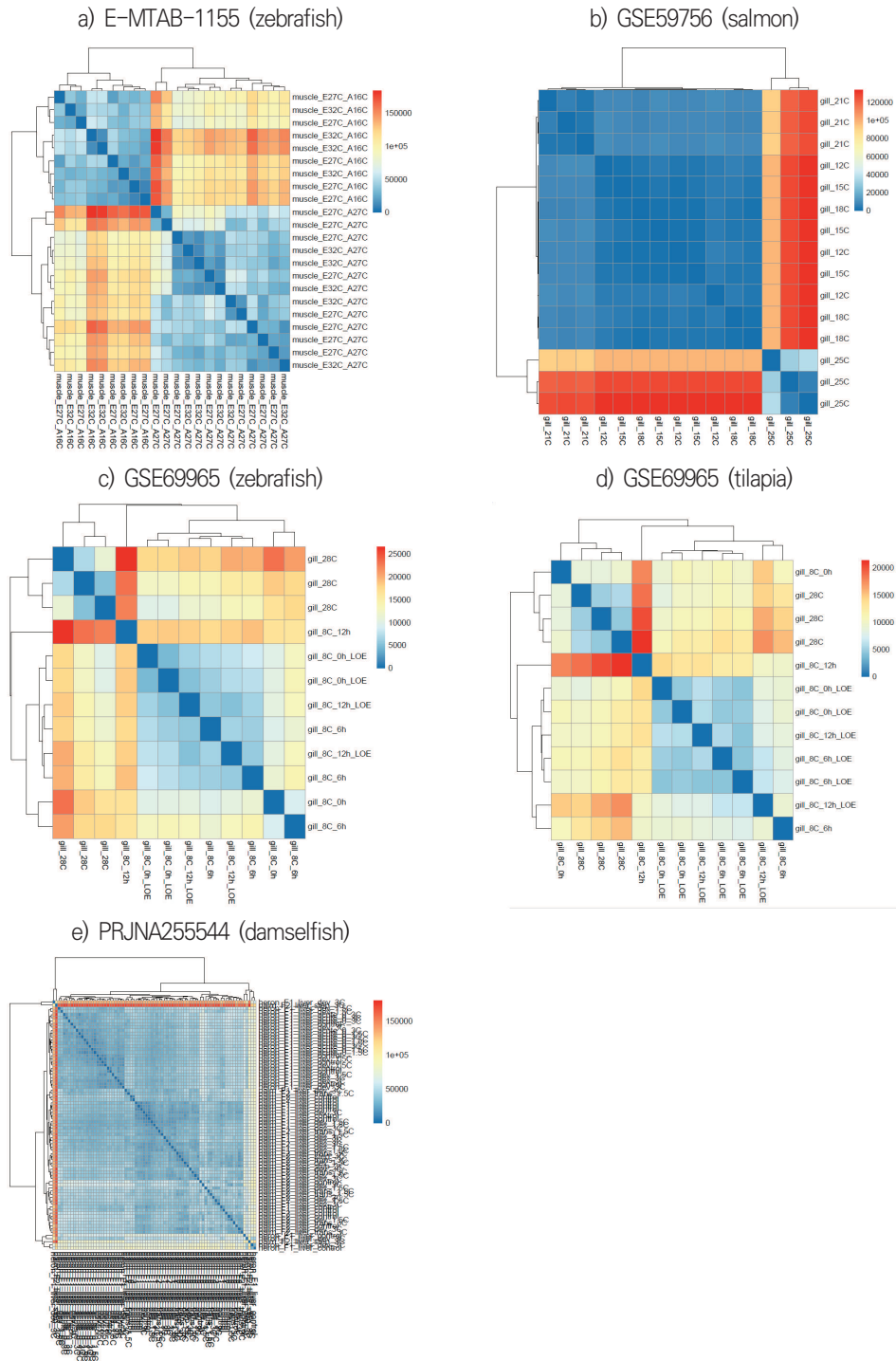


그림 4. 샘플간 distance matrix. 파란 색에 가까울수록 유사한 샘플, 빨간 색에 가까울수록 다른 샘플을 의미한다.

육안으로 관찰한 결과, Principal component analysis(그림 2)보다 Multidimensional scaling(그림 3)에서 샘플간 분리가 더 잘 일어났다. 데이터셋별로는 salmon(GSE59756), zebrafish(GSE69965), tilapia(GSE69965)를 대상으로한 실험에서 더 분리가 잘 일어났지만, 다른 zebrafish(E-MTAB-1155), damselfish(PRJNA255544) 데이터셋에서는 분리가 잘 되지 않았다(그림 2b-d, 3b-d). 그러나 E-MTAB-1155 dataset에서는 adult 시기의 온도(A16C과 A27C 샘플)에 따라 구분이 잘 되었고, PRJNA255544 dataset에서는 population에 따라(heron과 palm) 구분되었음을 볼 때(그림 2a, 2e), 새로운 feature 발굴 등의 추가적인 정밀 분석을 통한 샘플 분리가 가능할 것을 예상해 볼 수 있다. 샘플의 분리는 classification등의 후속분석의 성공 가능성을 알려주기도 하기 때문에, 지속적인 데이터의 특성 파악 및 새로운 feature 선택이 요구된다.

### 3.4 DEG 분석

차별발현유전자(differentially expressed genes)는 서로 다른 실험 조건의 전사체를 비교했을 때 통계적으로 유의미하게 발현량이 변화하는 유전자를 의미한다. 특정 환경에 따른 생체 반응을 이해하는데 결정적인 도움을 주는 유전자일 확률이 크기 때문에, 전사체 분석의 핵심적인 요소로 간주된다. 수집된 데이터셋으로부터 통계적으로 견고한 결과를 얻기 위해 샘플이 3개 이상인 condition을 선택하고 이들의 쌍방 비교로부터 차별발현유전자들을 구하였다.

표 6. 차별발현유전자 개수

Condition 1	Condition 2	#of DEGs
Dataset: PRJNA255544 (damselfish)		
heron_F1_control	heron_F1_dev_1.5C	1324
heron_F1_control	heron_F1_dev_3C	278
heron_F1_dev_1.5C	heron_F1_dev_3C	4060
palm_F1_control	palm_F1_dev_1.5C	449
palm_F1_control	palm_F1_dev_3C	276
palm_F1_control	palm_F2_control	2439
palm_F1_control	palm_F2_dev_1.5C	1382
palm_F1_control	palm_F2_dev_3C	1655
palm_F1_control	palm_F2_step_3C	2206
palm_F1_control	palm_F2_trans_1.5C	1992
palm_F1_control	palm_F2_trans_3C	2402
palm_F1_dev_1.5C	palm_F1_dev_3C	2
palm_F1_dev_1.5C	palm_F2_control	1264
palm_F1_dev_1.5C	palm_F2_dev_1.5C	539
palm_F1_dev_1.5C	palm_F2_dev_3C	598
palm_F1_dev_1.5C	palm_F2_step_3C	1123
palm_F1_dev_1.5C	palm_F2_trans_1.5C	655
palm_F1_dev_1.5C	palm_F2_trans_3C	1066
palm_F1_dev_3C	palm_F2_control	2006
palm_F1_dev_3C	palm_F2_dev_1.5C	401
palm_F1_dev_3C	palm_F2_dev_3C	423
palm_F1_dev_3C	palm_F2_step_3C	1272
palm_F1_dev_3C	palm_F2_trans_1.5C	749
palm_F1_dev_3C	palm_F2_trans_3C	800
palm_F2_control	palm_F2_dev_1.5C	381
palm_F2_control	palm_F2_dev_3C	192
palm_F2_control	palm_F2_step_3C	584
palm_F2_control	palm_F2_trans_1.5C	437
palm_F2_control	palm_F2_trans_3C	2253
palm_F2_dev_1.5C	palm_F2_dev_3C	5
palm_F2_dev_1.5C	palm_F2_step_3C	152
palm_F2_dev_1.5C	palm_F2_trans_1.5C	688

표 6. 차별발현유전자 개수(계속)

Condition 1	Condition 2	#of DEGs
palm_F2_dev_1.5C	palm_F2_trans_3C	926
palm_F2_dev_3C	palm_F2_step_3C	8
palm_F2_dev_3C	palm_F2_trans_1.5C	92
palm_F2_dev_3C	palm_F2_trans_3C	205
palm_F2_step_3C	palm_F2_trans_1.5C	36
palm_F2_step_3C	palm_F2_trans_3C	170
palm_F2_trans_1.5C	palm_F2_trans_3C	623
Dataset: E-MTAB-1155 (zebrafish)		
E27C_A16C	E27C_A27C	4562
E27C_A16C	E32C_A16C	902
E27C_A16C	E32C_A27C	5410
E27C_A27C	E32C_A16C	3275
E27C_A27C	E32C_A27C	13
E32C_A16C	E32C_A27C	13
Dataset: GSE59756 (salmon)		
12C	15C	0
12C	18C	443
12C	21C	2177
12C	25C	18462
15C	18C	217
15C	21C	1686
15C	25C	20080
18C	21C	555
18C	25C	21449
21C	25C	15847
Dataset: GSE69965 (zebrafish)		
28C	8C_0h_LOE	8178
28C	8C_12h_LOE	10224
28C	8C_6h	9624
8C_0h_LOE	8C_12h_LOE	180
8C_0h_LOE	8C_6h	351
8C_12h_LOE	8C_6h	292

표 6. 차별발현유전자 개수(계속)

Condition 1	Condition 2	#of DEGs
Dataset: GSE69965 (tilapia)		
28C	8C_0h_LOE	3642
28C	8C_12h_LOE	6992
28C	8C_6h_LOE	5932
8C_0h_LOE	8C_12h_LOE	219
8C_0h_LOE	8C_6h_LOE	92
8C_12h_LOE	8C_6h_LOE	239

표 6에서 보여지듯이 두 조건간 차별발현유전자 개수는 수개~수만개로 다양하다. 주로 loss of equilibrium(LOE)와 정상 조건 간의 차이에서 많은 수의 차별발현유전자들이 나오는데, 이는 생체 현상이 비정상적으로 되면서 대다수의 유전자 발현량에 영향을 끼치기 때문으로 추측된다. 또한 전사체의 경우 환경 변화에 빠르게 변화하기 때문에 발현량의 변동성이 큰 것으로 알려져 있고, 이 때문에 실제 환경 변화에 적응하기 위한 핵심적인 역할을 하지 않는 유전자들도 차별발현유전자에 포함된 것으로 추정된다.

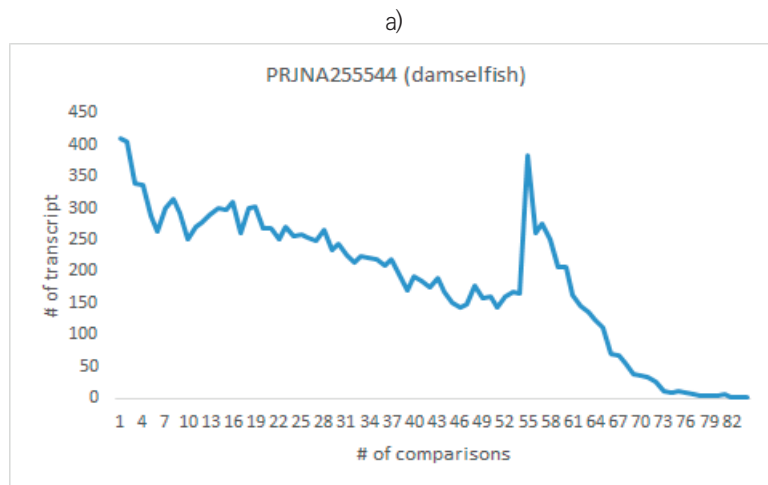


그림 5. 데이터셋별로 개별 DEG가 나타난 횟수

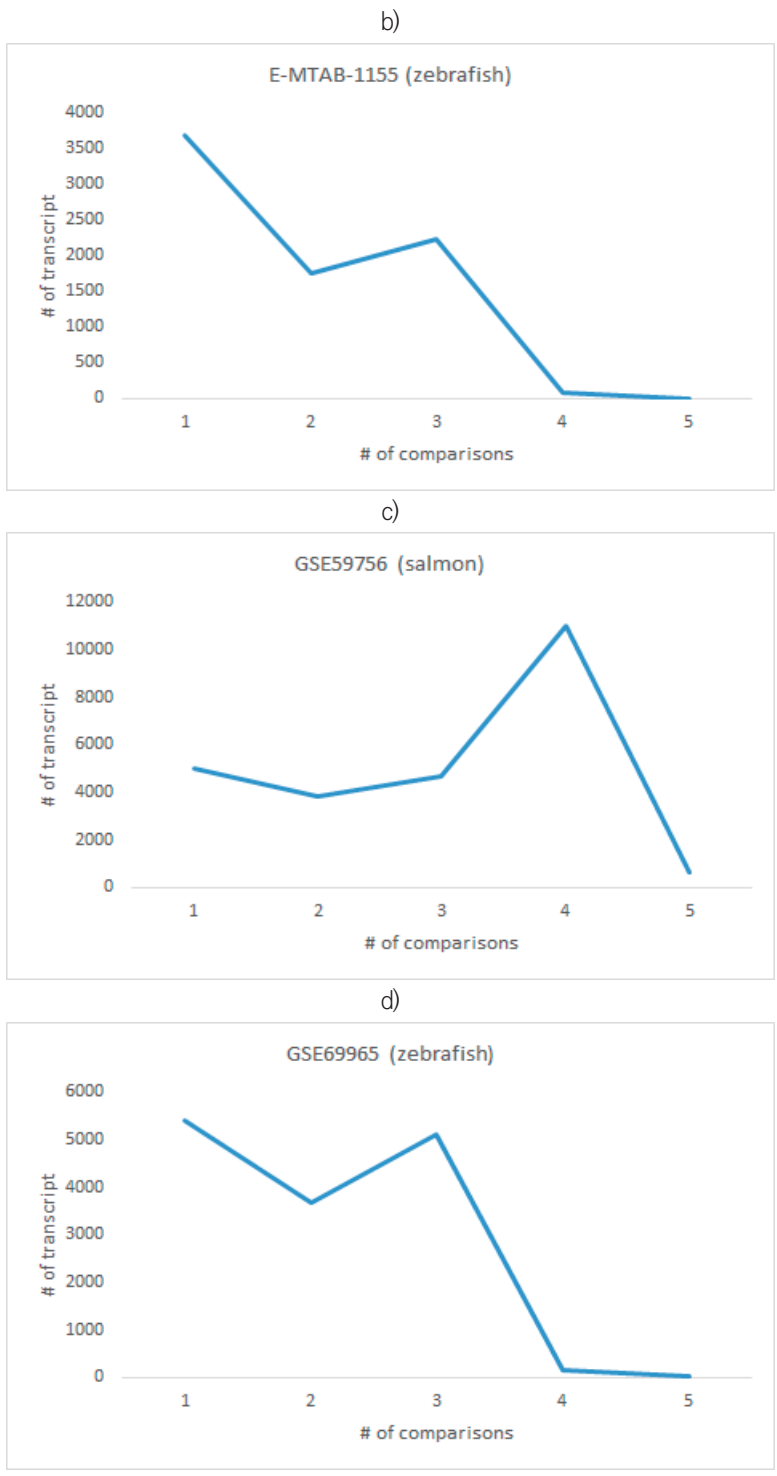


그림 5. 데이터셋별로 개별 DEG가 나타난 횟수(계속)

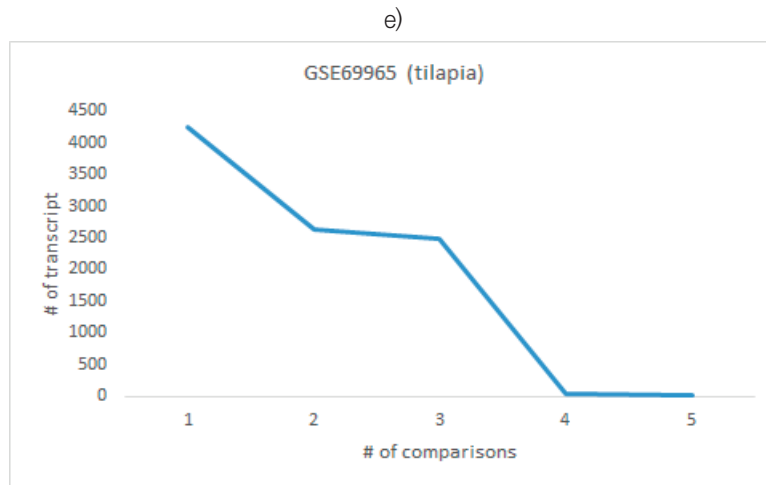


그림 5. 데이터셋별로 개별 DEG가 나타난 횟수(계속)

데이터셋 내의 샘플간 비교를 통해 차별발현유전자를 찾게 되는데, 이들이 실험 조건에서 어느 정도의 빈도로 차별적 발현 양상을 보이는지 알아보았다. 그림 5에서 보듯, 수백~수천개의 유전자들이 공통적으로 여러 실험 조건에 걸쳐서 발현량을 변화시킴을 알 수 있었다.

다른 데이터셋 간의 차별발현유전자의 유사성도 중요하다. 실험 조건의 차이가 크거나 생물종간의 진화적 거리 등이 멀수록, 차별발현유전자가 적을 것으로 예상되고, 환경/종-특이적인 유전자들의 역할이 커질 것으로 간주된다. 여러 데이터셋에서 공통적으로 나타나는 차별발현유전자들은 어류의 환경 변화 적응/실패에 핵심적인 기능을 하는 것으로 추정된다. 표 7에 나타난 것처럼, 두 데이터셋 간에 수천개의 공통적인 차별발현유전자들이 밝혀졌고, 이 중 281개의 차별발현유전자는 모든 5개의 데이터셋에서 공통적으로 나타났다(표 8). 어류의 온도 적응과 관련해서, 혈관 생성 및 확장, 콜레스테롤 등 영양분의 효율적인 전달과 이용을 가능케 하는 유전자들이 차별적으로 발현됨을 알려져 있다(Veilleux et al., 2015). 또한 온도 변화는 세포 입장에서 큰 스트레스 요인으로, 스트레스 반응과 관련된 산화, DNA 손상, 염증 반응을 억제해주는 유전자들의 발현이 크게 변화시킨다. 281개의 유전자들 중 어떠한 기능이 통계적으로 유의미하게 많이 나오는지 알아보기 위해 Gene Ontology를 이용한 enrichment 분석을 하였고, 이 중 혈관 생성(vasculogenesis), 활성산소(reactive oxygen species) 처리, 인슐린 반응성

과 관련된 기능이 특별히 많이 조절됨을 알 수 있었다(표 9). 이는 기존에 알려진 온난화 관련 반응과 크게 다르지 않음을 보여주는 것으로, 이들 차별발현유전자를 feature로 사용할 수 있는 당위성을 증명해준다. 추후 연구에서 이들의 정성적인 분석을 진행하여, 기온 변화에 핵심적인 역할을 하는 유전자들을 밝히고, 이들의 유전자마커로서의 가능성을 규명할 예정이다.

표 7. 데이터셋 간의 공통인 차별발현유전자수

Dataset 1	Dataset 2	공통 차별발현유전자 수
E-MTAB-1155 (zebrafish)	GSE59756 (salmon)	3,070
E-MTAB-1155 (zebrafish)	GSE69965 (zebrafish)	3,314
E-MTAB-1155 (zebrafish)	GSE69965 (tilapia)	1,748
E-MTAB-1155 (zebrafish)	PRJNA255544 (damselfish)	1,637
GSE59756 (salmon)	GSE69965 (zebrafish)	4,560
GSE59756 (salmon)	GSE69965 (tilapia)	3,211
GSE59756 (salmon)	PRJNA255544 (damselfish)	3,190
GSE69965 (zebrafish)	GSE69965 (tilapia)	2,708
GSE69965 (zebrafish)	PRJNA255544 (damselfish)	2,342
GSE69965 (tilapia)	PRJNA255544 (damselfish)	1,704

표 8. 모든 데이터셋에서 공통으로 나온 차별발현유전자와 기능 요약

Gene Names	Gene ontology
haus6	cell division
chaf1a	cell cycle; DNA repair
nusap1	mitotic cytokinesis
Map7d1	microtubule cytoskeleton organization
Bag6	covalent chromatin modification; intrinsic apoptotic signaling pathway in response to DNA damage by p53 class mediator; protein stabilization; regulation of apoptotic process
wdr37	maturation of 5.8S rRNA
sacm1lb	phosphatidylinositol dephosphorylation
chfr	cell division; mitotic cell cycle checkpoint
mettl2a	tRNA methylation
NLRP3	inflammatory response; positive regulation of transcription from RNA polymerase II promoter

표 8. 모든 데이터셋에서 공통으로 나온 차별발현유전자와 기능 요약(계속)

Gene names	Gene ontology
BHLHE40	negative regulation of transcription, DNA-templated; rhythmic process
Gls	glutamate biosynthetic process; regulation of respiratory gaseous exchange by neurological system process; suckling behavior
usp13	autophagy; cell proliferation; protein K63-linked deubiquitination
brd4	covalent chromatin modification; positive regulation of I-kappaB kinase/NF-kappaB signaling; regulation of inflammatory response
Hnrnpc	mRNA splicing, via spliceosome
NOP56	rRNA modification
CHD2	cellular response to DNA damage stimulus; muscle organ development;
CAPN9	digestion
BTAF1	negative regulation of chromatin binding
ABCC5	transmembrane transport
Kif1c	cytoskeleton-dependent intracellular transport
Map3k5	activation of MAPK activity; stress-activated MAPK cascade
SMO	multicellular organism development
apoeb	cholesterol biosynthetic process
PAPSS1	skeletal system development; sulfate assimilation
THAP12	negative regulation of cell proliferation; response to stress
EIF4G3	regulation of translational initiation
Tnk2	cell differentiation
RREB1	regulation of transcription, DNA-templated
NUAK1	cell adhesion
SYNCRIP	cellular response to interferon-gamma; RNA splicing
CLPX	ATP metabolic process; protein folding
UBXN7	post-translational protein modification
ABCA1	cellular response to cholesterol; response to nutrient
ZBED1	positive regulation of transcription from RNA polymerase II promoter
CYB5A	response to cadmium ion
ITGB2	cell adhesion; positive regulation of angiogenesis
RAF1	cell differentiation; multicellular organism development
CEACAM5	C-terminal protein lipidation; negative regulation of anoikis
GPI	angiogenesis; glucose homeostasis
Col4a2	angiogenesis; transcription, DNA-templated
Polr2a	cellular response to oxygen levels; transcription from RNA polymerase II promoter

표 8. 모든 데이터셋에서 공통으로 나온 차별발현유전자와 기능 요약(계속)

Gene names	Gene ontology
ALOX5	leukotriene biosynthetic process; lipoxin metabolic process [GO:2001300]; lipoxygenase pathway
CALM1	detection of calcium ion; regulation of cell communication by electrical coupling involved in cardiac conduction
EIF5A	mRNA transport; protein transport
Prf1	cytolysis; defense response to tumor cell
CEACAM1	angiogenesis; blood vessel development
GTF2F2	mRNA splicing, via spliceosome
Pde4b	cAMP catabolic process
MMP9	cellular response to reactive oxygen species; negative regulation of apoptotic process; positive regulation of release of cytochrome c from mitochondria
COL6A3	cell adhesion
CD22	cell adhesion
RPS6KB1	behavioral fear response; G1/S transition of mitotic cell cycle; response to heat
APC	cell cycle arrest; cellular response to DNA damage stimulus; insulin receptor signaling pathway
YY1	anterior/posterior pattern specification; cellular response to DNA damage stimulus
STOM	positive regulation of protein targeting to membrane
ATP6V1A	cellular response to increased oxygen levels
BTG1	negative regulation of cell proliferation; positive regulation of angiogenesis
HSPA4	chaperone-mediated protein complex assembly
MYH9	blood vessel endothelial cell migration
CHKA	lipid metabolic process; lipid transport
SREBF1	cellular response to fatty acid; cellular response to starvation; response to food
epd	cell-matrix adhesion
Cd9	cell adhesion
CSK	adaptive immune response; innate immune response
Nolc1	positive regulation of cell proliferation; response to osmotic stress
Crebbp	cellular response to hepatocyte growth factor stimulus; histone acetylation; positive regulation of CREB transcription factor activity;
IQGAP1	cellular response to epidermal growth factor stimulus
LIMS1	cell junction assembly; positive regulation of gene expression
CPT1A	cellular response to fatty acid; circadian rhythm; eating behavior; regulation of insulin secretion; regulation of lipid metabolic process
smc2	cell cycle

표 8. 모든 데이터셋에서 공통으로 나온 차별발현유전자와 기능 요약(계속)

Gene names	Gene ontology
Pa2g4	negative regulation of apoptotic process; rRNA processing; transcription, DNA-templated
JUN	transcription, DNA-templated
ADAR	adenosine to inosine editing; hematopoietic progenitor cell differentiation
PNP	nucleoside metabolic process
mcm2	cell cycle
SESN3	cellular response to amino acid starvation; cellular response to glucose starvation; glucose homeostasis; response to insulin
NLRP12	activation of cysteine-type endopeptidase activity involved in apoptotic process
YWHAE	cellular response to heat
ESRRG	positive regulation of transcription, DNA-templated
Gtf2b	positive regulation of core promoter binding
TRA2B	cellular response to glucose stimulus
RPS6KB1	apoptotic process
atp2a1	calcium ion transport; regulation of striated muscle contraction
PSPH	L-serine biosynthetic process; response to nutrient levels
EIF4G2	cell cycle arrest; cell death
SMAD3	transcription, DNA-templated
SRSF3	mRNA 3'-end processing
Vps45	protein transport
Dnajb9	ubiquitin-dependent ERAD pathway
Cry1	circadian rhythm; glucose homeostasis; lipid storage
DAB2	apoptotic process
EWSR1	regulation of transcription, DNA-templated
RHAG	ammonium transport; carbon dioxide transport
OGDH	2-oxoglutarate metabolic process; NADH metabolic process
CENPE	antigen processing and presentation of exogenous peptide antigen via MHC class II; regulation of mitotic metaphase/anaphase transition
MLLT1	negative regulation of protein kinase activity; transcription from RNA polymerase II promoter
ATP2A2	calcium ion transport
Col6a1	cellular response to amino acid stimulus
PTK2	angiogenesis; extracellular matrix organization; vasculogenesis
MEF2C	apoptotic process; blood vessel development; cellular response to glucose stimulus

표 8. 모든 데이터셋에서 공통으로 나온 차별발현유전자와 기능 요약(계속)

Gene names	Gene ontology
ZFP36L1	3'-UTR-mediated mRNA destabilization; apoptotic process; cellular response to hypoxia; cellular response to insulin stimulus; cellular response to salt stress; vasculogenesis
cnot7	transcription, DNA-templated
ccnyl1	regulation of cyclin-dependent protein serine/threonine kinase activity DNA integration
AHNAK	protein oligomerization; regulation of RNA splicing
TMOD4	actin filament organization; muscle contraction
ARHGEF5	actin cytoskeleton organization; positive regulation of apoptotic process
CDH17	adherens junction organization; transport
LCP2	cytokine secretion; immune response
GRB10	insulin receptor signaling pathway
KLF5	angiogenesis
RBM39	mRNA processing; RNA splicing
RRP1B	negative regulation of phosphatase activity; rRNA processing
SLMAP	muscle contraction; regulation of membrane depolarization during cardiac muscle cell action potential
BRD3	covalent chromatin modification; regulation of transcription from RNA polymerase II promoter
NECTIN1	cell-cell adhesion; immune response
RAB35	antigen processing and presentation; endosomal transport
CCNG2	cell cycle
cyp51	cholesterol biosynthetic process via 24,25-dihydrolanosterol
tmem106b	dendrite morphogenesis; lysosome localization
CYLD	homeostasis of number of cells; innate immune response
zmat1	intrinsic apoptotic signaling pathway by p53 class mediator
Ciart	circadian regulation of gene expression; transcription, DNA-templated
Pbxip1	cell differentiation; negative regulation of transcription, DNA-templated
Znf710	regulation of transcription, DNA-templated
ddx46	liver morphogenesis; regulation of gene expression
IFI44L	defense response to virus
hibch	valine catabolic process
NFYC	transcription, DNA-templated
nt5c2	adenosine metabolic process

표 8. 모든 데이터셋에서 공통으로 나온 차별발현유전자와 기능 요약(계속)

Gene names	Gene ontology
ABCC1	transmembrane transport
ZCCHC8	mRNA processing
U2SURP	RNA processing
TRIM35	apoptotic process; innate immune response
gnptab	carbohydrate phosphorylation; lysosome organization
lrmpl	mitotic spindle organization; mRNA transport
chac1	apoptotic process; glutathione catabolic process; response to unfolded protein
KIAA1217	embryonic skeletal system development
OBSCN	positive regulation of apoptotic process
RPRD2	snRNA transcription from RNA polymerase II promoter
ZNF318	meiotic cell cycle; transcription, DNA-templated
prrc1	activation of protein kinase A activity
Tmc7	ion transport
NT5C2	adenosine metabolic process
CNPPD1	regulation of cyclin-dependent protein serine/threonine kinase activity
DKC1	mRNA pseudouridine synthesis
EIF2S1	cellular response to amino acid starvation; translational initiation
Ddx5	alternative mRNA splicing, via spliceosome; androgen receptor signaling pathway; circadian rhythm
dhdh	D-xylose catabolic process
Vcl	adherens junction assembly; cell adhesion
ndfip1l	metal ion transport; negative regulation of gene expression; vascular transport
setd7	heart morphogenesis; transcription, DNA-templated
Trim39	apoptotic process; protein ubiquitination
sort1	transport
prmt6	DNA repair; transcription, DNA-templated
ccnf	cell cycle; protein ubiquitination
adm1b	proteasome assembly; ubiquitin-dependent protein catabolic process
arl8ba	cell cycle
rbbp4	cell cycle; transcription, DNA-templated
wdr61	positive regulation of transcription elongation from RNA polymerase II promoter
glrx5	cell redox homeostasis; hemopoiesis
fam213a	oxidation-reduction process
grtp1a	activation of GTPase activity

표 8. 모든 데이터셋에서 공통으로 나온 차별발현유전자와 기능 요약(계속)

Gene names	Gene ontology
rhpn2	signal transduction
USP53	action potential
USP54	protein deubiquitination
DHRS11	estrogen biosynthetic process
uba2	protein sumoylation
wtap	cell cycle; mRNA methylation
Fbxo11	cellular protein modification process
EXOSC4	DNA deamination; histone mRNA catabolic process
TTC21B	ventricular system development
dye	nuclear pore complex assembly
vcp	cellular response to DNA damage stimulus
vrk1	cell cycle; cell division
reep3	cell division
uap111	UDP-N-acetylglucosamine biosynthetic process
strip1	cortical actin cytoskeleton organization; heart contraction
mib1	blood vessel development
RASSF3	regulation of apoptotic process
Nelfa	negative regulation of transcription elongation from RNA polymerase II promoter
Csnk1g2	endocytosis; regulation of cell shape
Hlf	positive regulation of transcription from RNA polymerase II promoter; rhythmic process
Larp4	cytoskeleton organization
Tbc1d4	cellular response to insulin stimulus; intracellular protein transport
pol	DNA integration; transposition, DNA-mediated
ZZZ3	regulation of transcription, DNA-templated
ULK2	autophagy; response to starvation
Mk11	actin cytoskeleton organization
RASSF8	signal transduction
PER2	circadian rhythm; transcription, DNA-templated
Iqsec1	actin cytoskeleton organization
WDR20	protein deubiquitination
supt6h	chromatin maintenance; heart development
LMO7	protein polyubiquitination; regulation of cell adhesion
STAB2	angiogenesis; cell adhesion; regulation of blood coagulation

표 8. 모든 데이터셋에서 공통으로 나온 차별발현유전자와 기능 요약(계속)

Gene names	Gene ontology
TTN	cardiac muscle contraction
hsp90a.1	muscle organ development; protein folding; response to stress
IRF8	cellular response to lipopolysaccharide; transcription, DNA-templated
SOCS3	cytokine-mediated signaling pathway; negative regulation of apoptotic process; negative regulation of insulin receptor signaling pathway
fgfr1a	cardiac muscle tissue development multicellular organism development; transcription, DNA-templated
CAD	glutamine metabolic process hemolysis in other organism
Ppan	regulation of cell growth by extracellular stimulus
gjc1	cell communication
TEF	multicellular organism development; positive regulation of transcription, DNA-templated
RORB	cellular response to retinoic acid; rhythmic process
APBB2	cell cycle arrest; negative regulation of apoptotic process
SYTL4	calcium ion-regulated exocytosis of neurotransmitter
PLEKHG4B	regulation of Rho protein signal transduction
CSRNP1	apoptotic process; platelet-derived growth factor receptor signaling pathway; skeletal system morphogenesis
HNRNPAB	epithelial to mesenchymal transition; positive regulation of transcription, DNA-templated
Trim16	histone H3 acetylation; response to growth hormone
RIOK1	positive regulation of rRNA processing
UACA	apoptotic signaling pathway
Etv5	cellular response to oxidative stress
Rrs1	regulation of signal transduction by p53 class mediator; ribosomal large subunit export from nucleus
pola1	cell proliferation
Atp13a1	cellular calcium ion homeostasis
Tpcn1	calcium ion transport; positive regulation of autophagy
ADNP	cellular response to extracellular stimulus
ZNF644	regulation of transcription, DNA-templated
SCAF1	mRNA processing
SLC25A22	ion transport; L-glutamate transport; mitochondrial transport
smad1	BMP signaling pathway; embryonic pattern specification

표 8. 모든 데이터셋에서 공통으로 나온 차별발현유전자와 기능 요약(계속)

Gene names	Gene ontology
atnb233	potassium ion transport; sodium ion transport
ZFP64	regulation of transcription, DNA-templated
KIF13B	cytoskeleton-dependent intracellular transport
CUTC	copper ion homeostasis
SLTM	apoptotic process; transcription, DNA-templated
COMMD8	regulation of transcription, DNA-templated
TRMT1	tRNA modification
ZNF226	regulation of transcription, DNA-templated
tacc3	cell cycle
CISH	cytokine-mediated signaling pathway; protein ubiquitination
Xk	amino acid transport; cellular magnesium ion homeostasis
Pkp3	desmosome assembly; negative regulation of mRNA catabolic proces
Serinc1	membrane biogenesis
Nsdhl	cholesterol biosynthetic process
Fzr1	cell cycle; DNA repair
TBK1	I-kappaB kinase/NF-kappaB signaling
ACSL6	acyl-CoA metabolic process
ZMIZ1	artery morphogenesis; cell aging; vasculogenesis
TRAK1	anterograde axonal transport of mitochondrion
HEBP2	negative regulation of mitochondrial membrane potential
SQOR	sulfide oxidation, using sulfide:quinone oxidoreductase

표 9. Top 20 Biological Processes of Gene Ontology

GO ID	Biological Process
GO:0001570	vasculogenesis
GO:0007229	integrin-mediated signaling pathway
GO:0010594	regulation of endothelial cell migration
GO:0010694	positive regulation of alkaline phosphatase activity
GO:0030224	monocyte differentiation
GO:0031953	negative regulation of protein autophosphorylation
GO:0033762	response to glucagon
GO:0035026	leading edge cell differentiation
GO:0035994	response to muscle stretch
GO:0043484	regulation of RNA splicing
GO:0045577	regulation of B cell differentiation
GO:0045657	positive regulation of monocyte differentiation
GO:0045860	positive regulation of protein kinase activity
GO:0046627	negative regulation of insulin receptor signaling pathway
GO:0051894	positive regulation of focal adhesion assembly
GO:0051963	regulation of synapse assembly
GO:0060261	regulation of transcription initiation from RNA polymerase II promoter
GO:0071560	cellular response to transforming growth factor beta stimulus
GO:1901031	regulation of response to reactive oxygen species
GO:1903829	positive regulation of cellular protein localization

### 3.5 Classification System

본 연구를 비롯한 많은 오믹스 연구의 목적은 특정 질병 및 환경 변화에 중요한 역할을 하는, 따라서 진단 혹은 치료의 타겟이 되는 유전자를 밝히는 것이다. 육안으로 식별이 불가능한 상태에서 샘플의 유전자 자체의 서열 혹은 전사체의 발현량 등을 이용해 기존에 축적된 데이터와의 비교를 통해 두 가지 혹은 여러 가지 상태 중 어느 것에 가까운지를 측정하는 과정을 거쳐야 한다. 데이터간 비교 분석을 위한 Feature selection을 위해 다음의 과정을 거쳤다. 데이터셋 별로 여러 treatment간의 pairwise DEG를 구하여, 이들의 union set을 classification algorithm의 feature로 하였다. 이후 널리 사용되는 classifier인 support vector machine(svm) 과 naive Bayes classifier를 사용하여 전사체를 이용한 기온 변화 적응의 performance를 측정하였다. 데이터셋별로 training과 test set의 비율을 1:1로 고정하였다.

표 10. SVM classification. TP = true positive; FN = false negative; FP = false positive; TN = true negative

Dataset	TP	FN	FP	TN	Sensitivity	Specificity
E-MTAB-1155 (zebrafish)	0	6	0	5	1	0
GSE59756 (salmon)	3	0	4	0	0	1
GSE69965 (zebrafish)	4	0	2	0	0	1
GSE69965 (tilapia)	0	3	0	3	1	0
PRJNA255544 (damselfish)	28	0	11	0	0	1

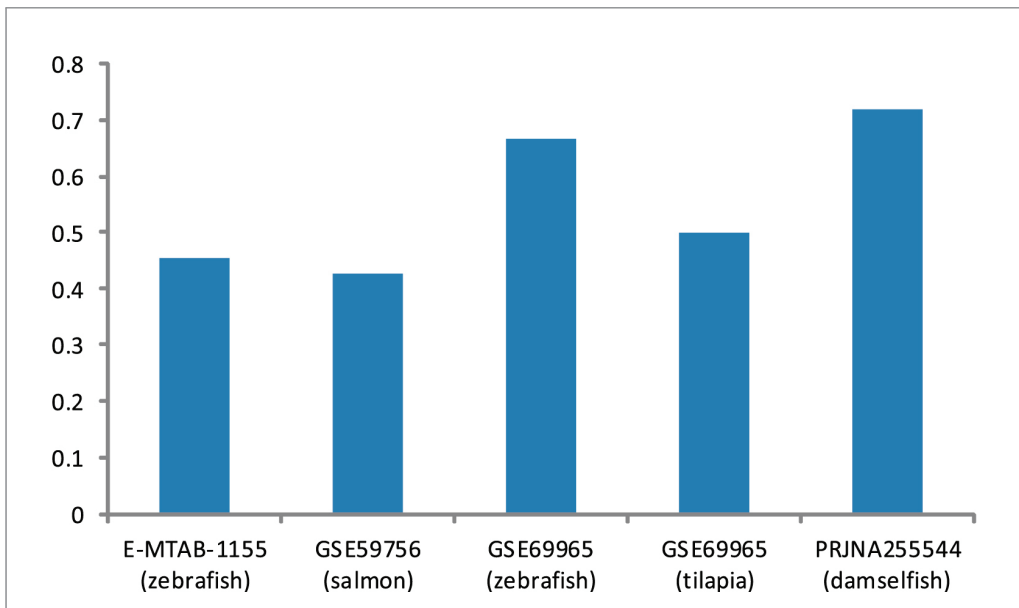


그림 6. SVM Classifier의 accuracy

표 11. Naive Bayes classification. TP = true positive; FN = false negative; FP = false positive; TN = true negative

Dataset	TP	FN	FP	TN	Sensitivity	Specificity
E-MTAB-1155 (zebrafish)	0	6	0	5	1	0
GSE59756 (salmon)	3	0	2	2	0.5	1
GSE69965 (zebrafish)	4	0	2	0	0	1
GSE69965 (tilapia)	3	0	0	3	1	1
PRJNA255544 (damselfish)	25	3	10	1	0.09091	0.89286

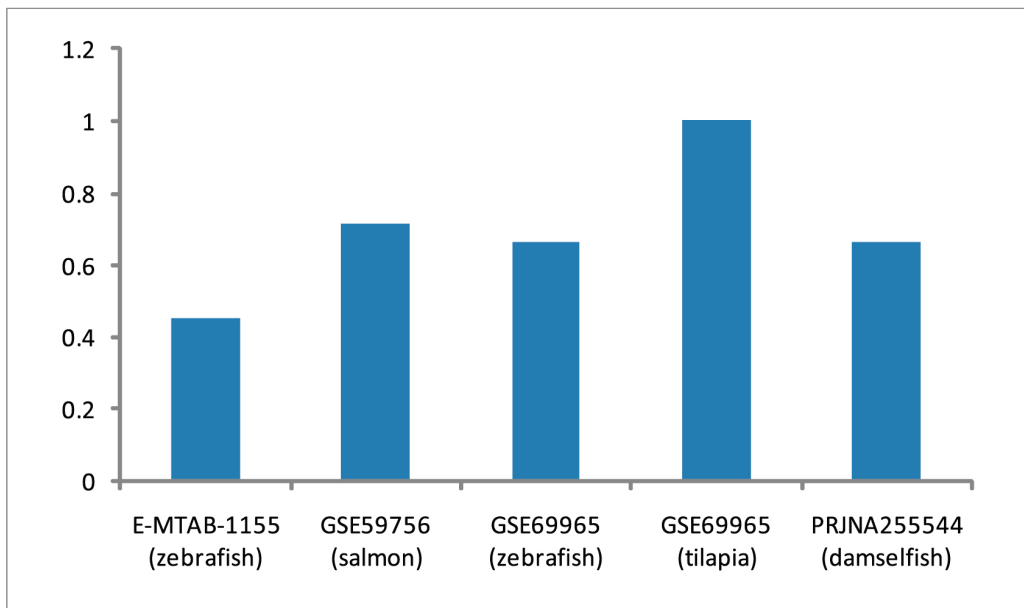


그림 7. Naive Bayes classifier의 accuracy

위에서 보듯, 데이터셋간 혹은 알고리즘 간의 분류 성능에 큰 차이를 보였다(표 10, 그림 6, 표 11, 그림 7). GSE69965(zebrafish)와 PRJNA255544(damselfish) 데이터셋의 경우 SVM에서 비교적 양호한 분류의 정확성(~70%)를 보였다. 반면 naive Bayes classifier를 이용했을 때는 E-MTAB-1155(zebrafish) 데이터셋을 제외한 나머지 4개 데이터셋에서 60%이상의 정확도를 보였다. 특히 GSE69965(tilapia) 데이터셋의 경우에는 높은 수치를 보였다. 양호한 분류 정확성에도 불구하고, 일반화에는 아직 어려움이 따른다. 우선, feature selection이 더욱 정밀해져야 한다. 이번 연구에서는 모든 실험 조건을 이용해 밝혀진 DEG들을 모두 feature로 사용하였지만, 이 중 subset이 더욱 좋은 정확도를 만들어낼 수 있다. 또한 데이터셋들의 사이즈가 PRJNA255544(damselfish)를 제외하고는 상당히 작기 때문에, training/test set 구축이 용이하지 않았다. 향후 추가될 데이터셋의 가공/분석을 통해 classifier의 성능을 증진시킬 예정이다.

## 4. 결론

가속화되는 기후변화와 기상이변은 식량 안보 및 생태계에 큰 영향을 끼친다. 현재까지 대부분의 연구는 특정 조건에서 개별 종의 소량 개체를 수년간의 생태 실험을 통해 그 현상을 관찰하는 단계에 있다. 그러나 개별 종을 일일이 관찰하는 것은 상당한 시간과 자본이 소요되며, 심지어 같은 종이더라도 집단에 따라 환경에 대한 반응 양상이 다르다는 것이 밝혀지고 있기 때문에, 새로운 접근법이 요구된다. 인간 질병 등 생명 연구가 많이 이루어진 분야에서는 갈수록 축적된 데이터를 효율적으로 가공이용할 수 있는 생물정보학의 중요성이 커지고, 실제 진단 마커 등의 시장 규모가 대폭 확대되고 있다. 생태 및 농수산 자원 분야에서도 데이터 수집, 가공, 해석을 가능케 하는 연구 및 파이프라인 구축이 시급하다.

본 연구에서는 수산자원을 대상으로 이루어진 기후변화 빅데이터 연구자료를 수집, 가공하여 데이터베이스화하였으며, 차별발현유전자의 분석, 및 자동분류시스템을 구축하였다. 향후 지속적으로 feature 선택 과정의 개선, 분류 알고리즘의 성능 개선, 데이터 축적 등을 이용하여 기후변화 연구에 핵심적인 역할을 할 수 있을 것으로 기대된다.

## REFERENCES

- Altschul, S.F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 1990: Basic local alignment search tool. *J Mol Biol*, **215**, 403–410.
- Bolger, A. M., M. Lohse, and B. Usadel, 2014: Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
- Boutet, E., D. Lieberherr, M. Tognolli, M. Schneider, and A. Bairoch, 2007: UniProtKB/Swiss-Prot. *Methods Mol Biol*, **406**, 89–112.
- Donelson, J., P. Munday, M. McCormick, and C. Pitcher, 2012: Rapid transgenerational acclimation of a tropical reef fish to climate change. *Nature Climate Change*, **2**, 30–32.
- Flato, G., and Coauthors, 2013: Evaluation of Climate Models. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Climate Change 2013*, **5**, 741–866.
- Grabherr, M. G., and Coauthors, 2011: Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*, **29**, 644–652.
- Haas, B. J., and Coauthors, 2013: De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*, **8**, 1494–1512.
- Hu, P., and Coauthors, 2016: Transcriptome comparison reveals a genetic network regulating the lower temperature limit in fish. *Sci Rep*, **6**, 28952.
- Langmead, B., and S. L. Salzberg, 2012: Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357–359.
- Lefevre, S., 2016: Are global warming and ocean acidification conspiring against marine ectotherms? A meta-analysis of the respiratory effects of elevated temperature, high CO<sub>2</sub> and their interaction. *Conserv Physiol*, **4**, cow009.
- Li, H., and Coauthors, 2009: The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Li, W., and A. Godzik, 2006: Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.
- Liu, S., and Coauthors, 2013: RNA-Seq reveals expression signatures of genes involved in oxygen transport, protein synthesis, folding, and degradation in response to heat stress in catfish. *Physiol Genomics*, **45**, 462–476.

- Love, M. I., W. Huber, and S. Anders, 2014: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.
- Scott, G. R., and I. A. Johnston, 2012: Temperature during embryonic development has persistent effects on thermal acclimation capacity in zebrafish. *Proc Natl Acad Sci U S A*, **109**, 14247–14252.
- Tomalty, K. M., M. H. Meek, M. R. Stephens, G. Rincon, N. A. Fangué, B. P. May, and M. R. Baerwald, 2015: Transcriptional Response to Acute Thermal Exposure in Juvenile Chinook Salmon Determined by RNAseq. *G3 (Bethesda)*, **5**, 1335–1349.
- Veilleux, H. D., and Coauthors, 2015: Molecular processes of transgenerational acclimation to a warming ocean. *Nature Climate Change*, **5**, 1074–1078.

**【연구자】**

유태우 선임연구원

연구보고서 2017-15

**해수 온난화에 의한 열대 수산자원 영향 연구**

---

I S B N 979-11-5698-205-0

발 행 2018년 3월

발 행 인 정 홍 상

발 행 처 APEC기후센터  
부산시 해운대구 센텀7로 12

제작·인쇄 경성문화사

---

이 보고서는 APEC기후센터 홈페이지(<http://www.apcc21.org/>)에서 이용하실 수 있습니다.  
또한 보고서에 실린 내용은 출처를 명시하면 자유롭게 인용할 수 있습니다. 단, 무단 전재 및  
복제를 금합니다.