DEVELOPING BIOMARKERS FOR MUSSEL LEUKEMIA AS TOOLS FOR ECOSYSTEM HEALTH MONITORING

by

EKATERINA VASSILENKO

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE STUDIES

(CHEMICAL AND BIOLOGICAL ENGINEERING)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

June 2012

Abstract

The overall goal of this project was to develop biomarkers for mussel leukemia, a sublethal endpoint used in population health monitoring. The most common method for leukemia assessment is hematocytology, which is labour intensive and subject to bias. In this thesis, new biomarkers based on DNA ploidy detection and genotyping, were developed and compared with hematocytology. These new biomarkers would allow for more sensitive and efficient monitoring of near-shore ecosystems. Two *Mytilus* species, *M. trossulus* and *M. edulis*, were tested. Samples were obtained from Hopkins and Horseshoe Bay beaches, a mussel farm (Island Scallops, BC) and from caged mussels of the same origins submerged at a monitoring site in Burrard Inlet and a reference site off the Sunshine Coast.

Three single nucleotide polymorphisms (SNPs) were detected within the coding region of p53 amplified from *M. trossulus* haemocyte cDNA, which were associated with leukemia. Many more polymorphic sites were found in *M. edulis*, some correlated with leukemia. Blocks in the p53 coding region sequences from late leukemic *M. edulis* were homologous with the *M. trossulus* p53 sequences, suggesting that hybridization may have contributed to increased disease susceptibility. Correlations between genotype and disease were not found in beach mussels or in either mussel species with early stages of leukemia.

DNA content flow cytometry patterns for *M. trossulus* haemocytes could distinguish healthy animals from diseased for all stages of leukemia, including early ones, where haemolymph contains mixtures of healthy and neoplastic cells. No strong association was observed between ploidy and leukemia in *M. edulis*. This new method for *M. trossulus* leukemia detection is recommended for monitoring of marine ecosystems exposed to multiple stressors, for which leukemia is a valuable endpoint together with other biomarkers required for efficient environmental management.

Preface

All parts of the research were performed by Ekaterina Vassilenko under the direct supervision and guidance of Dr. Susan Baldwin at the Department of Chemical and Biological Engineering. Assistance was provided in some specific areas by the following people: Chapter 2 (*M. trossulus* SNPs detection): Dr. Susan Baldwin, Dr. Annette Muttray (Research Associate in Dr. Baldwin's laboratory), Farida Bishay (Metro Vancouver) and Paul van Poppelen provided assistance with the fieldwork. Drs. Baldwin, Muttray and Dr. Patricia Schulte (Zoology Department at the University of British Columbia) contributed to the data analysis and editing of the written results. Chapter 3 (*M. edulis* SNP analysis): Dr. Susan Baldwin provided assistance with the data analysis and editing of the written results. Dr. Kermit Ritland (Faculty of Forestry) contributed to the application of population biology modeling to interpretation of the data. Chapter 3 (Ploidy analysis): Dr. Baldwin provided assistance with the data interpretation, on presentation and editing of the written results. Andy Johnson (Biomedical Research Institute, University of British Columbia) consulted on the performance of the ploidy experiments and analysis of the data.

Several undergraduate students working in Dr. Baldwin's laboratory assisted at some stages of the project. Caryn Liberman helped with some of the fieldwork, hemolymph sample collection and polymerase chain reaction (PCR) experiments during July-August 2008. Nancy Diao performed quantitative polymerase chain reaction (Q-PCR) experiments in July-August 2010.

A version of Chapter 2 was published in Mutation Research in 2010 (v. 701, p. 145-152). The title of the paper is "Variations in p53-like cDNA sequence are correlated with mussel leukemia: A potential molecular-level tool for biomonitoring", authors are Ekaterina I. Vassilenko, Annette F. Muttray, Patricia M. Schulte, and Susan A. Baldwin.

Material presented in Chapters 3 and 4 has been prepared for publication.

Ethical approval and certificates for my experiments were issued by the Animal Care Committee (University of British Columbia). The application numbers are A05-0057 and A08-0111.

Table of contents

List of tables

List of figures

List of abbreviations

µL	microlitre
ACL	alternate concentration limit
bp	base pairs
cDNA	complementary DNA
DNA	deoxyribonucleic acid
EDC	endocrine-disruption compounds
g	gram
HN	haemic neoplasia
HRM-PCR	high resolution melting polymerase chain reaction
HSPs	Heat Shock Protein
ITS	internal transcribed spacer
kb	kilobase
min	minutes
mL	millilitre
mM	millimolar
mRNA	messenger ribonucleic acid
nm	nanometer
^{o}C	degrees Celsius
PBS	phosphate buffered saline
PCR	polymerase chain reaction
qPCR	quantitative polymerase chain reaction
SNP	single nucleotide polymorphism

Acknowledgments

I would like to thank my supervisor Susan Baldwin and all members of the PhD committee, Patricia Schulte, Kermit Ritland and James Piret for their wise support and guidance.

I want to express my special appreciation to Annette Muttray, the former research associate in our group, for her constant and cheerful help and advice.

We would like to acknowledge Farida Bishay and Paul van Poppelen for providing assistance during the fieldwork. Dr. Rosemary Redfield is thanked for her comments on the data presentation in Chapter 2.

As well I want to acknowledge all helpful fellow graduate and undergraduate students in our laboratory, Jana Schmidtova, Parissa Mirjafari, Maryam Khoshnoodi, Caryn Liberman, Nancy Diao and Jocelyn Nelson.

Funding was generously provided by the Natural Sciences and Engineering Research Council of Canada Collaborative Research and Development Grant program (CRDPJ 349918) and Metro Vancouver (Head Office: 4330 Kingsway, Burnaby, B.C. V5H 4G8).

And finally I want to thank my husband, Evgeny Shapiro and my kids, Vladimir and Alexander, for inspiration and tireless support.

1 Introduction

The intertidal area is a very important element of the marine ecosystem, and it is intensively used as a food source and for recreation (FAO 1994). However, anthropogenic activity may have a significant negative impact on this ecosystem via overexploitation and contamination (Sherman 1994). Therefore, developing marine environmental effect monitoring programs that are able to identify environmental risks before adverse effects occur is a high priority. Two major approaches to environmental monitoring are detection of known contaminants (EPD 1993) and monitoring of biological responses of organisms (Environment_Canada 2012, Holt and Miller 2011). Each approach has a number of advantages and disadvantages, and implementation of both of them provides the most realistic picture of the environmental situation and its possible outcomes.

Unlike all chemical methods used for contaminant detection, biological responses to the presence of environmental contaminants provide information about ecological relevance. To make biomonitoring an efficient instrument of environmental management, seasonal and geographical variations of the biological responses need to be understood. In addition, biological responses and their outcomes vary depending on the biological species and their role in the ecosystem (Holt and Miller 2011).

Bivalve mollusks, in particular mussels, are among the key benthic species or "ecosystem engineers", providing a food source and facilitating a habitat for a wide range of other intertidal species (Crooks and Khim 1999, Gutierrez et al. 2003). Mussels modify the substrate and enrich the species richness (Borthagaray and Carranza 2007). Additional turbulent mixing created by mussels increases food supply to the near-bottom waters, thus changing the habitat conditions (Frechette' et al. 1989). Mussels are also of commercial interest as a significant food source for people in some areas (Canadian Aquaculture Industry Alliance, www.aquaculture.ca/files/species-mussels.php).

A number of mussel-based biomarkers have been developed that include responses at a wide range of biological levels, from sub-cellular (Dondero et al. 2006a) to population levels

(Widdows et al. 2002). Induction mechanism and ecological outcome for many of the responses have been investigated and well understood. This study is focused on mussel leukemia as a potential endpoint. Association of this disease with contaminants such as pesticides and oil products was observed in several studies (Brown 1980, Farley et al. 1991, Hillman 1993, Muttray et al. 2011), therefore mussel leukemia was proposed as a sublethal end point for near shore ecosystem monitoring programs. Molecular and physiological processes associated with mollusk leukemia were investigated and discussed in previous studies (Barber 2004, Kelley et al. 2001, Moore et al. 1991, Pariseau et al. 2011, Walker et al. 2006). However, there is no agreement among scientists on the etiology of the disease. Like all cancers, mussel leukemia is a complex disease, and there are most likely many factors that affect its initiation and progression. Nevertheless, the disease can serve as a non-specific and sensitive endpoint of population-level changes. In this study, mussels collected from several different beaches were caged together and exposed over several seasons to marine water at one urban and one remote location. To use mussel leukemia as an endpoint in monitoring programs, accurate and high-throughput methods to detect the disease are required. In this study, new disease markers based on genetic variations and aneuploidy were studied that can be included into environmental monitoring programs.

1.1 Biomonitoring

1.1.1 Challenges in biomonitoring

Sensitive and predictive methods for environmental monitoring are needed for early detection of deleterious anthropogenic effects. Biomonitoring is the science of estimating the ecological health of an area by studying sentinel organisms or indicator species living there (Cullen 1990). Biomonitoring can be applied to any ecological system, and it is often used to determine the effect of water quality on aquatic biota. Biomonitoring provides a way of measuring the effects of chemical contaminants and their mixtures on biological systems and its ecological outcome (Lam 2009). However, ecosystems are very complicated and they are exposed to various stressors, including complex mixtures of contaminants whose effects are not always simple to assess (Picado et al. 2007). In addition, many other confounding factors, such as season as well as spatial and genetic variations, make it difficult to attribute biological

responses to particular contaminants. In order to address these issues, an integrated multilevel approach is considered, where molecular changes are linked to biological consequences at all levels; individual, population and community (Picado et al. 2007).

There are three broad categories of biomonitoring (deZwart et al. 1995):

- Bioaccumulation monitoring: This involves measuring the accumulation of potentially toxic chemicals in different biological organisms and tissues. This is only informative if toxicity threshold levels of the chemicals are known, since chemical testing does not provide any information about the effects of the contaminants (Lam 2009). For example, measurement of the accumulation of organochlorine insecticide (Petrocelli et al. 1973) or metals (Chan 1988) in mollusk tissues does not provide information about the biological effects of these contaminants.
- Toxicity Assessment: Molecular and physiological responses of individual organisms towards toxicants are measured using bioassays. This approach is used to find contaminant concentrations at which specific biological endpoints are observed. For example, stress markers and morphologic characteristics are used to measure the effects of organic compounds on mollusk endocrine status (Gagné et al. 2007).
- Ecosystem monitoring: This measures how the ecosystem as a whole responds to environmental perturbations. This type of biomonitoring includes inventories of species diversity, density, and presence of indicator species.

Most biomonitoring programs encompass all of the above categories and activities follow the typical schedule below (Burga and Kratochwil 2001):

- An inventory of all living biota at all taxonomic levels is taken.
- Indicator species are identified and their tissues analyzed for chemical, physiological and other toxicity markers.
- Abiotic conditions are measured. In aquatic systems this includes chemical analysis of water and sediments as well as measuring of physical parameters such as temperature and turbidity, for example.

- Temporal and spatial variations of abiotic and biotic parameters are recorded. As many anthropogenic causes and natural sources of the variations as possible must be taken into account.

1.1.2 Biomarkers

The term “biomarker” was introduced and defined by the National Academy of Science (USA) as “a xenobiotically induced variation in cellular or biochemical components or processes, structures, or functions that is measurable in a biological system or sample.” (National_Research_Council 1987). In practice, biomarkers refer to the endpoints of biological tests, but there is no strong agreement among scientists on the use of the term. The term is sometimes used in relation to responses at the molecular or physiological levels, whereas others include responses at the organism or even population level (Lam 2009). The most valuable application of biomonitoring is providing an early warning of environmental and ecosystem changes (Tegler 2001). Therefore, biomarkers that provide information about significant biological effects are the most valuable. It is generally accepted that sub-organism level responses precede those that occur at higher levels of biological organization (Baldwin and Hochachka 1970, Gil and Pla 2001, Stegeman 1992). Thus, molecular, or physiological exposure biomarkers are of most interest as they can inform about changes taking place within an organism before irreversible population-level effects occur.

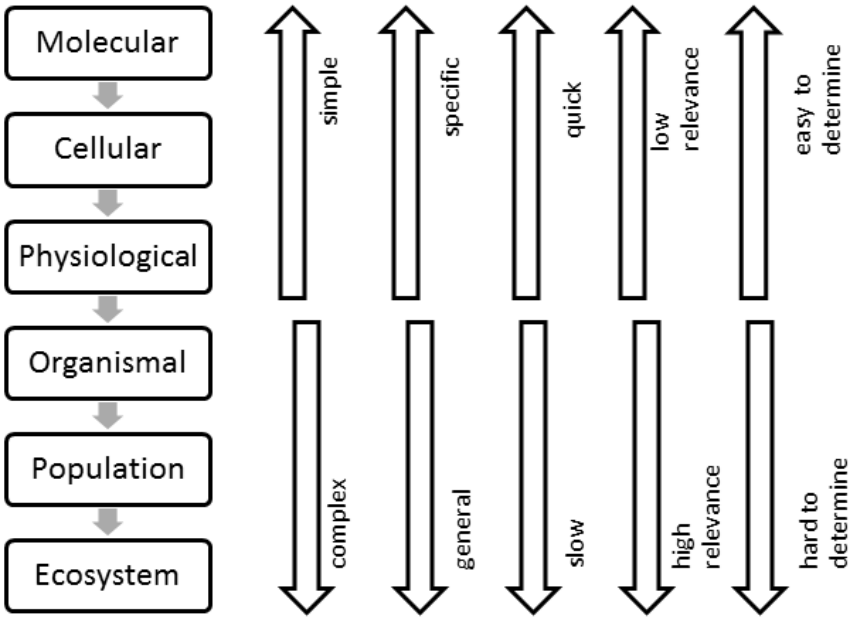


Figure 1.1. A comparison of the pros and cons of biomarkers at various biological organizational levels, from low (molecular) to high (population and ecosystem).

Multiple levels of biological organization are distinguished and advantages and disadvantages of using biomarkers at these levels are illustrated in Figure 1.1. (Connell et al. 1999, Fossi et al. 2000, Yeom and Adams 2007). Factors to consider when choosing biomarkers include the following.

- Complexity. Some responses that are easy to measure may arise from complex events. Examples are change in expression of a gene such as metallothionein and growth rate of the organism. Growth rate is very simple to assess, while gene expression measurement requires relatively sophisticated techniques and equipment. However growth rate is regulated by numerous factors through various pathways, which are sometimes difficult to identify, while regulation pathways for metallothionein are well understood (Cosson 2000{Rotchell, 2001 #697, Rotchell 2001). Therefore, the objective of the monitoring

program, such as the effect of a particular contaminant or ecosystem health, will dictate biomarker choice.

- Specificity. Biomarkers respond to stressors with various degrees of specificity. In general, specificity significantly decreases when going from molecular to ecosystem levels, although specificity might vary within one level. Most biomarkers respond to several contaminants (Lam 2009). Although general or non-specific biomarkers are efficient for detecting physiological stress, it is difficult or impossible to determine the specific stressor.
- Ecological Relevance. There are many enzymatic biomarkers available, and most have a strong correlation between their activity and exposure to a particular contaminant. However the question whether the enzyme activity is associated with harmful effects at the animal and population levels remains open (Gagnon and Holdway 2002, Lam 2009) and, therefore, the biomarker might have low ecological relevance. Some enzymatic biomarkers, such as lysosome activity, are not contaminant specific since many chemical, physical and environmental stressors are known to destabilize lysosomal membranes (Moore 1985). However they have ecological significance since they are linked to loss of key physiological processes. Physiological response-based biomarkers, such as growth, reproduction, or feeding are very popular, in particular for marine bivalve monitoring. These biomarkers are ecologically relevant since they provide an "integrated measure of organism well-being" or measure of "environmental quality", based on a wide range of functional attributes (Lam 2009). However they lack the specificity of molecular biomarkers. The challenge in selecting appropriate biomarkers is to find ones that link changes in critical physiological processes with specific stressors or contaminants.

1.1.3 Stress levels and biological responses

Organisms may respond to a particular stress through different mechanisms, depending on the length and intensity of exposure (Anestis et al. 2007, Wu 2005). Figure 1.2 illustrates the relationship between the response type and substance concentration (adapted from (Connell et al. 1999). It proposes that for essential compounds there is a range of concentrations between c1 and c2, where there is no harmful effect on the organism and for non-essential, potentially toxic, compounds there is a concentration, c2, below which there are

no toxic effects on the organism. For an essential substance, concentrations outside the c1-c2 concentration range cause biological responses, which may be reversible, non-reversible or lethal.

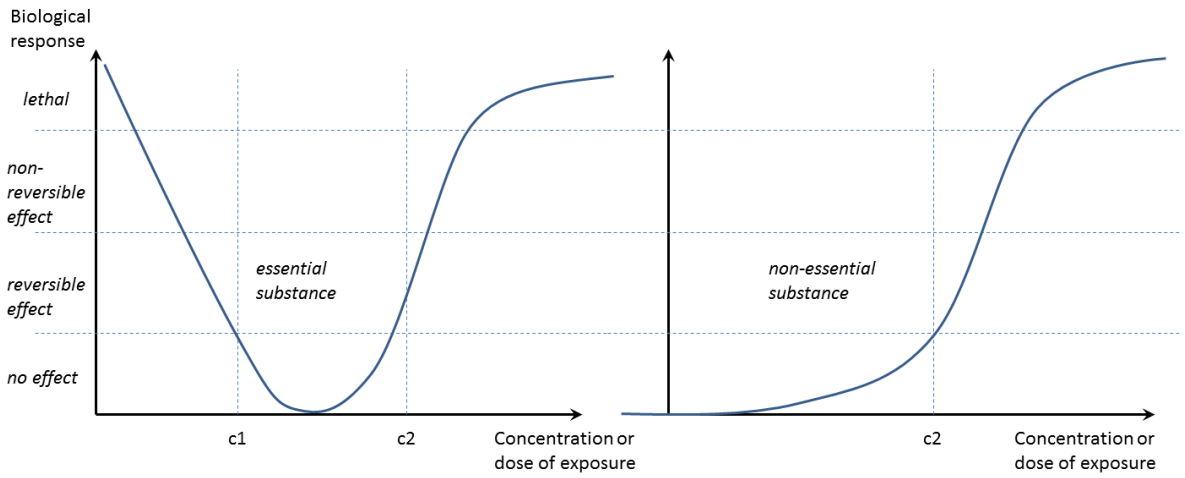


Figure 1.2. Biological responses to concentrations of substances essential to growth (such as nutrients) and non-essential (foreign). For most essential substances, there is a concentration range between c1 and c2 that provide "no stress" conditions. For non-essential and potentially toxic compounds there is concentration c2, below which there is no effect on the organism.

Reversible biological effects are useful for biomonitoring as they can provide information about trends of changing environmental conditions. Irreversible and lethal effects are informative only when stress is arising and not when it is decreasing or disappearing. Timing of the stress response, such as initiation, adaptation and recovery, depend on the species, endpoint and level of biological organization (Wu 2005).

1.1.4 Molecular level of the stress response

The fastest molecular responses to an environmental trigger might occur in an alive organism via protein modifications, such as phosphorylation, ubiquitination, and acetylation (Gao 2010, Polo and Jackson 2011, Spoel 2010). Slower responses may include quantitative changes of the amount of protein produced via variations at the translation or transcription levels (Gracey 2007, Schulte 2004, Storey 2004) A more profound response would result from

qualitative changes in the expressed protein, for example, where the alternate protein has the same basic function, but with new properties that yield different biochemical consequences. This might be realized by the expression of other isoforms or isozymes (Baldwin and Hochachka 1970, Gracey 2007, Schulte 2004). Switching from one gene to another coding for the homologues isozymes makes the measurement of the first gene expression level irrelevant and misleading (Fields 2012, Poly 1997). To ensure adequate stress detection, various possible qualitative changes in the response should be taken into account.

An example of reversible stress response is the ability of the organism to repair damage caused at the molecular level. For example, formation of DNA adducts in mussel gill tissue is associated with presence of polycyclic aromatic hydrocarbons (PAHs) in both field and laboratory exposures (Canova 1998, Pisoni 2004, Venier and Canova 1996). DNA integrity or amount of single- or double-strand DNA breaks correlates with some environmental contaminants. Both parameters, DNA adducts and DNA strand breaks, can be accurately measured by assays such as Tunel (Negoescu et al. 1996) or comet (Machella et al. 2006). However the dose-related increase of adducts and amount of strand breaks disappear with time due to repair mechanism (Ching et al. 2001, Lam 2009). Another reversible stress response is the upregulation of the Heat Shock Protein (HSPs) superfamily. They become upregulated immediately as the stress is applied, but return to the normal level after removing the stress (Almoguera et al. 2002, Feder 1999). Therefore, molecular or subcellular responses might not lead to significant biological effects at the organism or population levels. Moreover, detecting a biomarker such as DNA adducts would indicate presence of a stressor, however absence of a response can mean either absence of the stressor or an efficient repair system (Lam 2009). Overall, both field validation and understanding of the molecular mechanisms involved with responses to various stress types and levels over time are crucial for obtaining the stressor relevance of biomarkers.

Bioavailability or the mode of entry of toxins into the organism should also be taken into account when selecting biomarkers (Higashi 2000, Moore et al. 1998). For example, respiratory surfaces experience a higher intake and response to contaminants as compared to the rest of the body surface for the same exposure dose. The fate of the toxin in organism is important as this determines the type of the organism response (VanEgmond 2004). The toxin

might be eliminated through secretion, transformation into another substance, or sequestration in tissues, which dictates the marker choice.

Biological relevance of the molecular biomarkers dictates classification into several categories such as markers of exposure, markers of the biological response and markers of susceptibility (Strickland and Groopman 1995). Markers of the exposure, such as metabolites and DNA adducts (Castano-Vinyals et al. 2004) indicates the dose of the toxic agent, response markers reflect biological response (Bresler et al. 2003, Snyder 2001) and markers of the susceptibility inform about the individual sensitivity to the toxic agent (DelMazo 2006, Thier et al. 2003).

Since many molecular level responses occur soon after exposure to a stressor, many molecular biomarkers of the exposure can be used as early warning indicators before observable damage at the higher biological levels occur. Molecular biomarkers have many advantages, discussed above, such as relevance to the stressor and biological outcome. Modern molecular biology tools can be used to analyze physiological stress or adaptive responses at many different levels, including signal transduction, transcription, translation, kinetic and allosteric controls, post-translational modification, subcellular localization, and protein degradation (Storey 2004). Nevertheless, there are caveats associated with using molecular level responses for assessing population health. For example, changes at the molecular level can vary hugely between individuals even though they are exposed to the same environmental stressors due to their differences in health condition, life history or genetic predisposition (Casanueva et al. 2012, Császár et al. 2009). Confounding factors, such as age, season of the year, location or even weather might affect the response and can interfere with biomarker sensitivity (Buesching et al. 2009, Picard and Schulte 2004, Wada et al. 2009).

Another drawback to using molecular level biomarkers is that they can be of a transient nature. For example, gene expression might return to levels similar to those in unstressed cells, even under prolonged stress (Almoguera et al. 2002, Wu 2005). Sometimes, heritable regulatory changes can result in long-term adaptation and evolutionary changes. This happens due to the high number of genetic variations in regulatory DNA elements (Carroll et al. 2004, López-Maury et al. 2008). Thus, genetic biomarkers can provide biased information

under situations of prolonged stress and, therefore, alternatives should be considered in the prolonged stress case.

The usefulness of molecular-level biomarkers can be significantly increased by measuring several of them at the same time. One way of doing this is to use DNA microarrays that allow for measurement of transcriptional responses associated with a wide range of stressors (Gracey 2007, Storey 2004). For marine bivalves, frequently employed as an indicator species in marine monitoring programs, several microarrays were developed. They include an expressed sequence tags (EST) – based microarray ((Venier et al. 2006, Venier 2009) MytArray, 1758 ESTs;) and a low density oligo-array ((Dondero et al. 2006b), Mytox chip, 24 mussel genes). MytArray includes a large number of sequences, but only 50% of them were putatively identified and associated with a limited number of pathways. These pathways may not be relevant to the endpoints or responses of interest, which is one disadvantage of EST-based microarrays, in particular of those developed for non-model organisms, such as mussels. For many genes on the Mytox Chip, their relevance to types of stresses and molecular pathways are known, although the number of genes currently identified for mussels is low, which is the limitation for this particular type of array. Other gene sequencing projects and microarray developments are underway (Craft et al. 2010, Place et al. 2012), which will increase the ability to apply genetics to environmental monitoring with mussels. From microarray data, it is possible to deduce or hypothesize the mechanisms of stress responses, which can guide the development of assays simpler than DNA microarrays.

1.1.5 Cancer biomarkers and their application in biomonitoring

Carcinogenesis is being studied almost exclusively in mammalian systems. However high incidences of neoplasia have been reported for other species such as fish (Baumann 2010) and mollusks (Barber 2004). Many chemicals, such as pesticides, chlordane and oil products, are known to be responsible for formation of mutations and tumours (Ashby and Tennant 1991, Baumann 1998, Benfenati 2009, Hillman 1993, Lowe and Moore 1978, Muttray et al. 2012, St-Jean et al. 2005). Most mutagens and carcinogens are strong electrophilic alkylating or acylating agents with various stressed heterocyclic rings (Beranek 1990, Fishbein 1980). Alternatively, they produce hydroxyl or oxygen radicals (Kanno et al.

2011, Kasai 1991). A number of carcinogens need to be metabolically activated to become harmful (Connell et al. 1999). The initiation of carcinogenesis is often attributed to DNA strand breaks and the formation of DNA adducts (Bernstein 2009, Hoeijmakers 2009, La 1996). However, the biological significance and ecological implications of carcinogens are not fully understood due to the complexity of the mechanisms involved and the time lag between the exposure and the formation of a detectable tumour (Armenian 1987, BRER 2006, Fraser 2011, Kikuchi 2004). As well, cells have sophisticated and flexible repair mechanisms, so low doses of carcinogens might not lead to any observable effect within a short period (BRER 2006), pp.65-90). After exposure to mutagenic conditions, cells with altered DNA sequences may continue to replicate and become detectable even a long time after the exposure. When abnormal cells undergo programmed cell death, the apoptosis-related biomarkers can be used as indicators of a genotoxic response (Ainsworth et al. 2011, Lant and Storey 2010). Formation of a tumor can be considered a useful non-specific biomarker, which definitely has a biological effect at organismal and population levels.

Genetic variation-based biomarkers are most useful when the genes have central functions in critical tumor pathways. A very good example is the tumour-suppressor p53 gene, a critical transcription factor that initiates cell-cycle regulation responses . The p53 gene was discovered more than three decades ago (Crawford 1983, Lane 1992, Lane and Crawford 1979) and intensively studied since then due to its critical role in cell regulation (Chen 2000, Collins et al. 1997, Giaccia and Kastan 1998, Giaccia and Kastan 2006, Shaw 1996).

1.1.6 P53-family genes

The vertebrate p53 family is coded by three genes: p53, p63 and p73 (Arrowsmith 1999, Arrowsmith 2003, Benard et al. 2003, Cesková 2003). These genes, like many other transcription factors, have a modular structure that consists of highly conserved domains, interrupted by variable regions (Figure 1.3). Highly conserved domains are shown in color. The transactivation domain (TAD) is specific for transcription co-regulators. The proline-rich domain (PXXP) is involved in signal transduction. The DNA binding domain (DBD) is a highly conserved region for sequence specific DNA binding. The oligomerization domain (OD) is responsible for the formation of the active p53 tetramer. Finally, the nuclear localization signal (NLS) domain is involved in localizing p53 in the cell nucleus

(Arrowsmith 2003, Benard et al. 2003, Liang and Clarke 2001). Highly conserved C-terminus domain with HOMO and SAM motifs was detected in p63/73 isoforms and it participates in tetramerization process (Chi et al. 1999, Jessen-Eller et al. 2002, Muttray et al. 2007).

A p63-like sequence is proposed to be the most ancient. It evolved in unicellular Choanozoans, (Rutkowski et al. 2010), where either one p53 or both p53 and p63/73 are present in Flagellates in various species (Nedelcu and Tan 2007). Several independent duplication events of the original p63-like gene occurred in different animal lineages, such as warms, insects and vertebrates (Rutkowski et al. 2010). The appearance of the p53 analogous genes predated the multicellular organisms, and, therefore, cancers. Consequently, tumour suppressor function of the 53 gene was not the primary function of the family (Fernandes 2008, Lu et al. 2009).

The p63 and p73 genes have functional blocks similar to p53 such as the TAD and DBD, although there are several features that make them different from the p53 gene (De Laurenzi 1998, Fillippovich 2001, Kaghad et al. 1997). Both p63 and p73 genes have an additional promoter, which gives rise to N terminal truncated (ΔN) isoforms lacking the TA domain (Yang et al. 1998, Yang 2000), (Figure 1.3, ΔNp73). As well, both genes have an additional sterile alpha motif (SAM), which is involved in protein-protein interactions (Arrowsmith 2003, Benard et al. 2003, Crawford 1983, Joerger and Fersht 2010). The two promoters and the possibility of alternative splicing at both N terminus and C terminus regions give rise to at least six possible isoforms per gene (De Laurenzi 1998, Fillippovich 2001, Kaghad et al. 1997, Yang 2000). TA forms of both p63 and p73 are known to be involved in cell differentiation, cell cycle regulation and apoptosis in a manner similar to p53, whereas ΔN isoforms are known to act as antagonists to all TA forms of the family and are considered to be oncogenes (Benard et al. 2003).

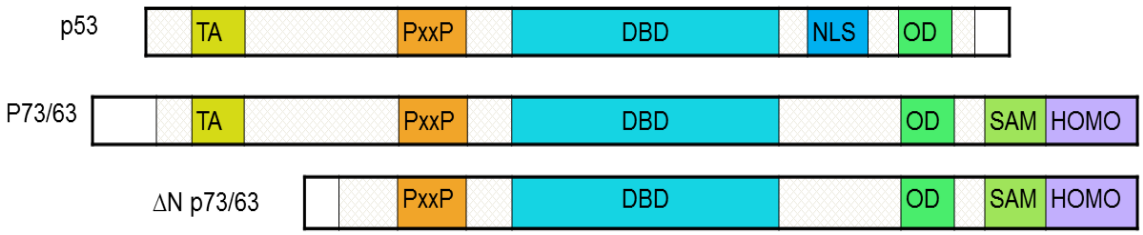


Figure 1.3. Human p53 family protein domains: TAD-transactivation domain, PxxP-proline-rich domain, DBD-DNA binding domain, OD-oligomerization domain, NLS –nuclear localization, SAM-sterile alpha motif and HOMO-homodimerization domain.

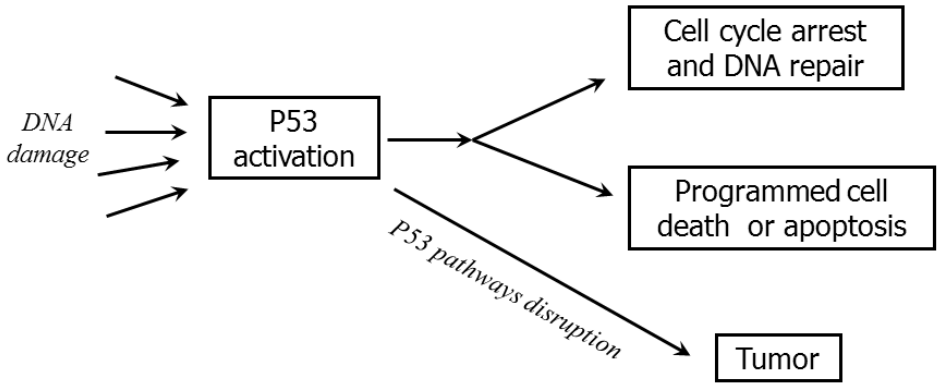


Figure 1.4 . p53 activation and function in response to DNA damage.

Expression of p53 is modulated by a wide variety of signals from the external and internal cell environment (Giaccia and Kastan 2006, Soussi 2012). One of the most important processes regulated by the p53 gene is cellular stress response, in particular that associated with DNA damage, which acts via regulation of the cell cycle and cell proliferation or apoptosis induction (Greenblatt et al. 1994). The simplified scheme of this p53 stress response, presented in Figure 1.4, illustrates three alternative scenarios: 1) cell cycle arrest and initiation of repair mechanisms to maintain cells; 2) elimination of those cells due to DNA damage and 3) uncontrolled cell replication, which can lead to carcinogenesis.

1.1.7 Mammalian p53 Mutations

P53 is mutated in more than 60% of human cancers, especially in cancers linked to environmental carcinogens. Some of these mutations have been shown to result in dysfunctional p53 proteins thus contributing to the deregulation of cell proliferation and the progression of cancer (Kato et al. 2003). Almost all non-viral tumours in humans, regardless of tissue of origin and including hematological cancers, exhibit point mutations in the coding region of the p53 gene, particularly in the highly conserved DNA binding domain (Cetin-Atalay and Ozturk 2000, Greenblatt et al. 1994). Because p53 plays such an important role in cancer, much work has been done characterizing genetic mutations and their implications for protein expression and function. These mutations can be used as potential biomarkers for early indications of changes that precede tumorigenesis. As such, genetic mutations can be useful in environmental monitoring for detecting sub-lethal endpoints. In addition, ΔN isoforms, which act as p53 antagonists, have been proposed as molecular markers of tumours in mammals (Becker et al. 2006).

A database is maintained of the distribution of mutation frequencies compiled from all current studies on p53 mutations in human cancer (Figure 1.5).

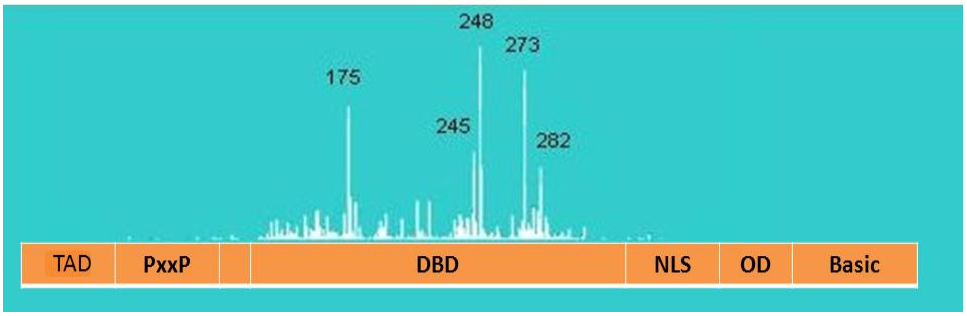


Figure 1.5. Somatic mutations in the coding region of human p53 gene (codon numbers, IARC TP53 Mutation Database, R15 release, November 2010, http://www-p53.iarc.fr/GraphSTAT.asp?TypeGraph=CodonDistribution).

There are five dominant hotspots (at codons 175, 245, 248, 273, and 282) common for both somatic and germ-line mutations. They are located mostly in the core DNA-binding region. Various cancers have different characteristic mutation hotspots, specific for tissue and cancer type and created by the influence of exogenous carcinogens (Kouidou et al. 2006).

Since p53 protein functions as a tetramer transcription factor, heterozygous mutations can modulate the function of wild-type p53. This dominant-negative inhibition was detected in many cancer types (Joerger and Fersht 2007). For example, heterozygous mutations were shown to cause severe defects in p53-dependent apoptosis in embryonic stem cells (Vries et al. 2002).

A number of cancers, such as smoker's lung cancer, prostate cancer, cancers of bone and connective tissues are associated with silent mutations (Kouidou et al. 2006, Strauss 2000). Silent mutations do not change amino acid sequence, however a nucleotide polymorphism (variations in nucleotide sequence) can have several consequences. It might cause a different secondary structure of the mRNA molecule, affecting its interaction with cellular components, which, in turn, would affect RNA processing, transport, stability and expression (Shabalina et al. 2006, Shen et al. 1999). Silent mutations can induce alterations in the splicing process, affecting the structure of the resulting protein (Chao et al. 2001). Silent mutations, leading to variations in codon usage frequency, can affect both protein expression rate and secondary structure of protein fragment downstream of the mutation (Cortazzo et al. 2002). Protein conformation changes affect the interaction of p53 with DNA or other proteins. Most of the p53 mutants found in cancer cells exhibit a variety of local structural variations, while the overall protein structure remains preserved (Joerger and Fersht 2007).

At the DNA level, silent substitutions may create additional sites of methylation, or eliminate them. Methylation is known to play a crucial role in the regulation of the cell differentiation (Latchman 2002). Under-methylation leads to the formation of a more open chromatin structure, which affects gene expression (Greenblatt et al. 1994). In the human p53 gene, all CpG sites (adjacent C and G if going from 5' end to 3' end) on both DNA strands have a methylated cytosine, which makes the methylation pattern tissue independent (Kouidou et al. 2005, Kouidou et al. 2006, Tornaletti and Pfeifer 1995). Moreover, non-CpG methylation has been reported recently, which occurs in CC-rich and CCWGG motifs. DNA from cancerous tissues was more heavily methylated at non-CpG cytosines, compared to the non-cancerous tissues (Kouidou et al. 2005). In addition to the well-known epigenetic effects, methylation provides the possibility of non-reparable C-T transitions via de-amination of methylated cytosine, thus creating additional hotspots of mutations (Toledo and Wahl 2006).

It is known that 80% of p53 mutations are associated with sites of methylation, although the causative role of mutations is not always proven. Due to its central role in cell cycle regulation and the strong possibility of carcinogenesis when dysfunctional, p53 was selected as a candidate biomarker for neoplasia in mollusks as will be described in detail later in the next section.

1.2 Marine biomonitoring

This thesis deals specifically with the application of biomarkers to monitoring of ecosystem health in the near-shore marine environment, which may be exposed to pollutants from many different sources such as:

- municipal sewage discharge
- discharges from boats
- agricultural run-off
- surface runoff from polluted areas
- industrial wastes

Since the near shore environment receives effluents from a wide variety of sources, monitoring and managing the effects of contaminants is complicated. Only a limited number of the many chemicals released to the environment (US EPA, Toxic Substances Control Act, "Chemicals of Concern" list) are being measured at any particular analysis event. As well, many chemicals of concern are toxic at very low concentrations, where they are difficult to measure. Some compounds are only toxic in combination with other chemical compounds (Kreiling et al. 2005). In these cases, toxic effects can only be detected by monitoring biological responses. Therefore, molecular biomarkers can be very useful for detecting the biological responses to complex environmental stressors of various intensities.

1.2.1 Mussels as indicator organisms

Aquatic invertebrates have the longest history of use as sentinel species in biomonitoring programs (Viarengo et al. 2007). In 1975, a "Mussel Watch" program was proposed "to assess the spatial and temporal trends in chemical contamination in estuarine and coastal areas of North America" (Goldberg 1975). Following this initiative in the U.S.A., many other countries, such as The United Kingdom, France, Canada, India, The Soviet Union and South Africa established similar Mussel Watch programs (Cantillo 1991, Rosenberg and Resh 1993). Use of bivalves, mussels in particular, can be explained by their many advantages as indicator organisms:

- Mussels dominate rocky coastal and estuarine communities and play a key role in the ecosystem (Crooks and Khim 1999, Gutierrez et al. 2003). As well, they have a wide geographical distribution (Gosling 1992).
- Being sessile, they assess the contamination of a particular area.
- Mussels are relatively tolerant, but not insensitive, to a wide range of environmental conditions (Anestis et al. 2007).
- Being suspension-feeders, they concentrate many chemicals in their tissues, making them much easier to detect (Cho et al. 2004, Roper et al. 1996, Salazar and Salazar 1996).
- Mussel populations are relatively stable and large, so they can provide statistically significant data (Comesana et al. 1999, Korrida et al. 2010, McGrorty and Goss-Custard 1993).
- Mollusks can be maintained in cages at sites of interest, where populations would not normally grow (St-Jean et al. 2003). This feature provides an additional possibility to standardize measurements.

Due to these advantages, Metro Vancouver (MV, formerly the Greater Vancouver Regional District, GVRD) is testing two species of blue mussel, *Mytilus edulis* and *M. trossulus* for their use in environment monitoring programs in the Burrard Inlet. These two species co-exist and hybridize on both European and Canadian Atlantic coasts, whereas on the Pacific coast only *M. trossulus* are found (Gosling 2003). Although the two species are very similar morphologically, they behave differently. *M. edulis* grow faster and reach larger sizes, while *M. trossulus* has a higher mortality rate and shorter life span (Innes et al. 2005). Interspecies

fertilization is limited by gamete incompatibility however is sufficient to maintain a low frequency of hybrids in natural populations (Toro et al. 2004). Survivability of the hybrids of the two species is lower than either pure species (Innes et al. 2005). *M. edulis*, as well as other *Mytilus* species found in Europe, have a lower susceptibility to mussel leukemia than *M. trossulus* (Barber 2004, Bower 1989, Bower 2006). Given these differences between *M. trossulus* and *M. edulis*, the results of a marine health monitoring program may be different depending on which species is deployed. Therefore, both species were used in the MV pilot project and both were analyzed and compared in this thesis.

Common measurements performed in environmental monitoring with mussels are bioaccumulation and specific responses to known pollutants (Table 1.1), such as cytochrome P450 induction and metallothionein induction (Peters et al. 1999, Sarkar et al. 2006). As yet, genetic biomarkers are not included on the list in Table 1.1 since they are still being developed.

Table 1.1. Examples of biomarkers included in biomonitoring programs utilizing bivalves.

Cytochemical methods	**Biochemical methods**
Lysosome membrane stability	Catalase activity
Lysosomal lipofuscin content	Malondialdehyde
Lysosomal neutral lipid	Total Oxidant Scavenging Capacity
CaATPase activity	Acetylcholinesterase activity
Peroxisome proliferation	Metallothioneins
Micronuclei	Glutathione-*S*-transferase in hemolymph
DNA damage	
Lysosome / cytoplasm ratio	
Growth rate	
Reproduction	
Mortality	

1.2.2 Bivalve leukemia as a sub-lethal endpoint for biomonitoring

Bivalve leukemia is a hemolymph cell disorder that has been registered for at least 15 different species of bivalves (House et al. 1998) and was described first by Farley for *Mytilus trossulus* from Yaquana Bay (Oregon, USA) and Sequim Bay (WA, USA) in 1969 (Brooks and Elston 1991, Farley 1969). The disease is also known as leukemia (McGladdery et al. 2001), haematopoietic neoplasia (Smolowitz and Reinisch 1993), disseminated sarcoma, disseminated neoplasia, epizootic sarcoma or transmissible sarcoma (Farley et al. 1991). This is a complex disorder, characterized by proliferation of non-adhesive mitotic cells with a high nucleus-to-cytoplasm ratio (Ciocan and Rotchell 2005, Gosling 2003, Kelley et al. 2001) and increased cellular DNA content of 3.8 to 7.7 times the diploid level (Elston et al. 1990). The disease was classified as leukemia because of continuously dividing neoplastic cells, which penetrate various tissues such as connective tissue, gonads, mantle and foot (Barker et al. 1997). At late stages of the disease, the hemolymph contains abnormal cells in high numbers ($6x10^8$ /mL, while the normal amount is $1\text{-}6x10^6$/mL) (Kelley et al. 2001), and beyond a certain stage, death of the organism occurs. The disease can contribute to mortalities of *M. trossulus* up to 90% over the annual season. There is no agreement among investigators on the seasonal variations in the prevalence of leukemia. The increase of the disease prevalence during the fall and winter months was reported (Leavitt et al. 1990, Rasmussen 1986), while other authors detected significant prevalence year-round (Brooks and Elston 1991, Farley et al. 1991, McGladdery et al. 2001).

Since 1978, much data has been collected addressing possible correlations between environmental pollutants and leukemia in mollusks. In surveys conducted with *Mya arenaria*, leukemia was detected at sites contaminated with jet fuel (Brown 1980), chlordane (Farley et al. 1991), polychlorinated biphenyls (PCBs), heavy metals, industrial wastes (Brown 1980, Farley et al. 1991, Reinisch et al. 1984, Yevich and Barszcz 1978). Higher leukemia prevalence (up to 70%) in the local bivalve populations were found at sites contaminated with PAH, chlordane, pesticides and cadmium (Barber 2004, Brown 1978, Hillman 1993, Lowe and Moore 1978, Van Beneden 1994). Studies of soft shell clams (*Mya arenaria*) from Sydney mines (NS), Kitimaat Arm, (BC) and along the east coast of the USA (sites containing high levels of anthropogenic contaminants) revealed high incidences of the disease

(McGladdery et al. 2001, Van Beneden 1994). Moreover, in some areas of Maine and Florida (USA) significantly increased levels of clam tumours in parallel with higher human mortality rates due to cancers were detected (McGladdery et al. 2001, Van Beneden 1997). In contrast, surveys conducted in Boston Harbor (MA, USA) on clams (Smolowitz and Leavitt 1996) and in France on oysters (Balouet et al. 1986) found no correlations between disease level and total contamination or oil pollution. On the other hand, one study failed to find a relationship between tissue content of environmental contaminants, including PAHs, PCBs and heavy metals, and the prevalence of HN in the mussels *Mytilus trossulus* in Puget Sound, Washington (Krishnakumar et al. 1999). Thus, it has not been shown conclusively that environmental pollutants play a key role in the prevalence of leukemia. The possibility that other confounding factors, such as season and transmission could also be involved and interrelation between these factors requires further investigation. To help identify the causes of mussel leukemia more accurate detection methods are needed especially for the early stages of the disease.

A recent study of leukemia in clams on Prince Edward Island (PEI, Canada) found a correlation between the incidence of late leukemia and proximity to potato farms that apply pesticides and fertilizers (Muttray et al. 2012) This study used a combination of methods of leukemia detection namely immunocytochemistry, which was confirmed by p53 family genes expression analysis. The immunocytochemistry done in this study used antibodies that were developed to discriminate between leukemic and normal haemocytes in *M. arenaria* (McGladdery et al. 2001, Smolowitz et al. 1989). To support the immunoassay results, expression of p53 family genes was employed, since previous studies have shown that expression of these genes was altered in leukemic haemocytes compared to normal (Muttray et al. 2010, Muttray et al. 2008). Leukemia detection methods and large amount of sites assessed make this study unique and the results strongly suggest that clam leukemia level correlates with the presence of water contaminants.

Since several monitoring experiments have found that this disease is associated with water contaminants, leukemia has been adopted as one of the endpoints used for environmental monitoring in Mussel Watch programs (Barber 2004, Van Beneden 1994). For a number of reasons, such as challenges in their reproduction and the requirement for special

skills and equipment, leukemia-specific antibodies are not available for wide use. Therefore, markers for the detection of all disease stages, especially early the stages, remain undeveloped and require further investigation.

1.2.3 Mussel haematocytology

Analysis of hemolymph composition based on the morphology of the cells is the main method for detecting leukemia (Farley 1976, Farley et al. 1991, Moore et al. 1991). Bivalve haemocytes are classified as granulocytes or agranulocytes (Calisi et al. 2008, Dyrynda et al. 1997, Hine 1999, Pipe et al. 1997, Ratcliffe and Rowley 1981) Immunological analysis and Percoll density gradient centrifugation (Dyrynda et al. 1997) have also separated out these morphologically different sub-populations of haemocytes, which have been shown to be functionally different. Granulocytes comprise the majority of hemolymph cells and are associated with phagocytosis and oxygen free radical generation. All cells that are large (7-12 µm) with conspicuous granules in their cytoplasm, a low nucleus to cytoplasm ratio and large spread pseudopodia when attached to the microscope slide, are classified as normal cells or granulocytes (Barber 2004, Bower 1989, Ratcliffe and Rowley 1981). Agranulocytes are generally smaller (4-6 µm) than granulocytes and do not have spread pseudopodia but much smaller protrusions. In addition, they are characterized by absence of granules and low cytoplasm volume. The roles of agranulocytes cells are not clear, although hyalinocytes (a type of agranulocytes) display characteristics of undifferentiated cells (Carballal et al. 1997).

Neoplastic cells, in advanced stages of the leukemia, are easy to detect since they are morphologically very distinct. They are round and smooth due to their inability to adhere or spread and they have very high nucleus to cytoplasm ratios (Barber 2004, Bower 2006, Elston et al. 1988, Smolowitz et al. 1989). The cell interior is characterized by mitotic features and sometimes several nucleoli. However, leukemia is difficult to recognize at the early stages since some neoplastic cells share morphological similarities with normal cells, as demonstrated in studies with antibodies developed against leukemic cells (McGladdery et al. 2001, Smolowitz et al. 1989).

Hematological analysis in mollusks is complicated by the presence of cell agglomerates or aggregates, in which it is difficult or impossible to identify and count cells.

Mollusk haemocytes often form clumps during collection of the haemolymph and sometimes anticoagulants are added to prevent clump formation and to facilitate visualization of the cells (Chen and Bayne 1995). Cell aggregates are believed to be formed as a result of the host response, phagocytosis and/or encapsulation response (Franchini et al. 1995, Gosling 2003, Jayaraj et al. 2009). According to Elston and co-authors, presence of aggregates in mollusk hemolymph can be associated with the regression of the leukemia (Elston et al. 1988). Although this issue has not been discussed or studied by other authors.

1.2.4 Mollusk p53 family

Analogues of well-known tumour suppressor p53 family have been identified in several mollusk species (Kelley et al. 2001, Muttray et al. 2005, Stephens et al. 2001). The structure of the gene in mollusks is similar to the vertebrate p63/p73 gene, while in fly *Drosophila melanogaster* (Ollmann et al. 2000) and the nematode *Caenorhabditis elegans* (Schumacher et al. 2001) the gene has p53-like structure. Exposure to organic contaminants, such as Benzopirene, caused upregulation of the p53 gene in the most of analyzed mussel tissues (Banni 2009). Development of leukemia is known to be associated with upregulation of some p53 family members in neoplastic haemocytes, at both protein and mRNA levels (Kelley et al. 2001, Muttray et al. 2010).

The domain structures of mussel p53, p63/73 and ΔNp73 mRNAs are similar to those of other taxa, including mammals (Muttray et al. 2005). Similarities in p53 coding sequences between distinct taxa vary depending on the region. For example, the transactivation region in *Mytilus* is 75% similar to that in *H. sapiens*, whereas the mussel DNA binding region similarity to *H. sapiens* varies from 63-81% depending on the domain (Muttray et al. 2005).

Mollusk p53-like and p63/73-like homologs have nearly identical sequence in their core DBD region (Cox et al. 2003, Kelley et al. 2001, Muttray et al. 2005, Muttray et al. 2007). From the analysis of p53 isoform haplotypes and preliminary Southern blotting experiments it was concluded that all mollusk p53 family members are probably the products of one gene (Belyi et al. 2010, Lu et al. 2009, Stifanic et al. 2009). Presence of the alternative promoter, the tandem alternative polyadenylation sites and complicated intron-exon structure provide possibilities to express all members of this diverse protein family from one gene copy (Muttray et al. 2007).

Expression of N-terminal TA- and ΔN isoforms was observed for clams and mussels (Kelley et al. 2001, Muttray et al. 2007). Upregulation of the ΔN isoforms is known to be associated with oncogenesis and, thus, these have been proposed as a biomarker for some human cancers (Zaika et al. 2002). Alternative splicing and 3'UTR variations give rise to the isoforms different at the C-terminus (Muttray et al. 2008). Alternate polyadenylation is well known in the differential expression of isoforms in a tissue- and stage-dependent manner during differentiation (Latchman 2002).

In mammals, the p53 family is expressed from three distinct genes and in human cancers p53 is often mutated, while p63 and p73 remain intact (Ponassi et al. 2006). If there is only one gene for the mollusk p53-like family, then mutations in the exons that are shared between all isoforms will affect the whole family. Conversely, mutations in the exons specific for only some isoforms would affect only some of the expressed p53 family genes. For instance, mutations in the TA domain would not cause any changes on delta-N isoforms, while mutations in SAM-domain would affect only p63/73 and leave p53 intact.

Significant variations in the expression of the p53-like family members were detected in mollusks with leukemia. In particular, mRNA levels of p53 and ΔNp63/73 mRNA were significantly higher ($p < 0.0001$) in neoplastic versus normal haemocytes. As well, p53-like and p73-like proteins, which are normally localized in the nucleus, were detected exclusively in the cytoplasm of leukemic haemocytes of clam, *Mya arenaria*, where they are thought to be inactive (Kelley et al. 2001). Barker et al. (1997) suggested that p53 was mutated in the 6^{th} exon in leukemic haemocytes, which corresponds to the evolutionary highly conserved coding region of the p53 sequence. Stephens at al. (2001) detected up-regulation of a p73-like protein in the end-stage leukemic clams, while in normal animals p73-like protein was not detected at all with an antibody specific to the highly conserved DBD of *Mya arenaria* p53 family. Although, this antibody cannot distinguish between TA and delta-N isoforms. Remarkably, the level of p53-like protein appeared to be the same between normal and leukemic animals (Kelley et al. 2001). However, a previous study revealed differential expression of p53 family members in leukemic and normal haemocytes of *Mytilus trossulus* at the mRNA level, where isoform-specific quantitative real-time reverse transcription PCR analysis showed significant

increases of p53-like and ΔN p63/73-like mRNA in the leukemic haemocytes (Muttray et al. 2008).

Human ΔN isoforms are known to suppress p53 activity, thereby promoting tumorigenesis (Bourdon et al. 2005, Zaika et al. 2002)**.** It is not yet known whether the mollusk p53 family members interact in the same manner, although similarity in their sequence structure suggests they may do so. Increased levels of p53 mRNA will not necessarily correspond to increased protein levels. Factors such as low mRNA stability, inefficiency of translation or protein instability can eliminate the effect of transcriptional up-regulation and, as a result, lead to low protein levels. Indeed, as it was already mentioned, p53 protein level was not increased in leukemic *Mya arenaria*, although it is not known if the same observation would be made in *Mytilus trossulus*.(Kelley et al. 2001). Variations at the RNA level, e.g. variations in mRNA that affect its secondary structure or stability, might counter-act the up-regulation of transcription and lead to inefficiently translated or dysfunctional p53 (Shabalina et al. 2006, Shen et al. 1999). Therefore, it was hypothesized that silent mutations in p53-like mRNA could have a causative effect in the development of leukemia in mussels.

1.2.5 Abnormal haemocyte ploidy

Genetic biomarkers have a number of pitfalls as were discussed earlier in the Section 1.1. Morphological or physiological changes, especially ones that can be measured in a non-biased and high-throughput way, may be useful as early-warning indicators of deleterious changes and may be more reliable than genetic biomarkers. The haemocyte morphology approach is a manual method that requires highly trained personnel and is laborious and biased (St-Jean et al. 2003). In addition, it is very problematic to diagnose the early stages of leukemia due to the ambiguous appearance of some haemocytes and low amount of neoplastic cells.

Another event that is associated with the formation of neoplastic cells is the accumulation of abnormal ploidy in circulating haemocytes. Several studies have shown that mollusk leukemia is associated with abnormal haemocyte ploidy in several bivalve species, including clams (Delaporte et al. 2008, Siah et al. 2008), cockle (da Silva et al. 2005) and mussels (Bihari et al. 2003, Gonzalez-Tizon et al. 2000, Moore et al. 1991, Smolarz et al.

2005). Cancerous cells are known to undergo various chromosomal rearrangements such as amplifications, deletions and translocations, which results in their being aneuploid rather than polyploid (Storchova and Pellman 2004). As in human cancers, the ploidy of mollusk cancer cells varies between different species, cancer types and individuals (Delaporte et al. 2008). The normal *Mytilus trossulus* diploid karyotype (2N) constitutes of 28 chromosomes, whereas the chromosome content in neoplastic cells varies from 48 to 92 (Gonzalez-Tizon et al. 2000). Moore and co-authors described two distinct types of leukemia in *Mytilus* based on their DNA content, namely the pentaploid and tetraploid forms, with a prevalence of 66% for the pentaploid form, 29% for the tetraploid form, and 5% exhibiting both disease forms simultaneously (Moore et al. 1991). In contrast, detailed karyotype analysis showed that ploidy was not the same in each diseased mussel individual (Gonzalez-Tizon et al. 2000).

Ploidy may play a direct role in carcinogenesis since accumulation of polyploidy promotes genetic instability and carcinogenesis, as it was shown for human epithelial cells (Roh et al. 2008). Due to these associations of ploidy and cancer, it may be a useful biomarker for the disease. Nevertheless, there are some confounding factors. Often there is no standard or reference aneuploidy type as individuals with a similar health status might vary in ploidy. Also, the question of whether aneuploidy causes the cancer or is it a consequence of it has not yet been solved (Storchova and Pellman 2004).

1.3 Objectives

The overall goal of this work is to develop efficient methods for mussel leukemia detection and to compare the leukemia prevalence in mussel populations exposed to different environmental conditions. Genetic biomarkers were targeted due to the increasing interest in the use of high-throughput genomics screening for environmental effects monitoring. To this end, the coding region of the p53 gene was screened for sequence variations and their association with the disease. This gene was targeted based on evidence for its involvement in HN shown in previous studies. The hypothesis was that there might be variations in the p53 gene that cause post-transcriptional deregulation of p53 expression and function. Two mussel species, *Mytilus trossulus* and *M. edulis*, were studied since both are being considered for environmental effects monitoring in the Burrard Inlet and their genetic dispositions may differ.

A cell physiology-based biomarker was investigated as an alternative to a genetic biomarker. DNA ploidy was chosen since this is a feature of neoplastic cells in many bivalves and had been suggested previously as a potential biomarker for mollusk leukemia. A flow cytometry method for screening mussel haemocytes was developed and used for statistical comparison of the disease incidences in several populations.

Both the genetic and DNA ploidy methods were compared with the standard visual-microscopy based technique for diagnosing HN.

The specific research questions were:

1. Are there sequence variations in the *Mytilus trossulus* p53-coding region that are associated with leukemia and which can be used as disease biomarkers?
2. Are there similar or other sequence variations in the *M. edulis* p53-coding region and does it matter which species is used in environmental monitoring programs?
3. Can polyploidy be utilized as a marker for population health for both *Mytilus* species?
4. Do we get the same results when comparing the health of different mussel populations using the two different methods: haemocytological diagnosis of mussel leukemia and detection of polyploidy?

Chapters 2 and 3 describe the results of variations in the mussel p53 gene and their associations with leukemia in two *Mytilus* species, *M. trossulus* (Chapter 2) and *M. edulis* (Chapter 3). Chapter 4 presents the development of a ploidy-based leukemia biomarker and its application for mussel population health monitoring. Finally, the Conclusions and Future Work Chapters discuss the potential outcome of the current study for marine environmental monitoring and mussel aquaculture.

2 Variations in p53-like cDNA sequence and their correlation with mussel leukemia in *Mytilus trossulus.*

2.1 Scope

Several bivalve species, including mussels (*Mytilus* spp.) and clams (*Mya* spp.) are susceptible to a leukemia-like disease that has been known to decimate whole populations (Brooks and Elston 1991, Farley 1969). Previous studies of molecular processes associated with late stages of this leukemia have implicated analogs of the p53 tumour suppressor protein family in disease etiology (Kelley et al. 2001). Synonymous single nucleotide polymorphisms (SNPs) in the coding region sequence of p53-like cDNA from *Mytilus trossulus* (blue mussel) were detected that were located at positions 182, 392 and 821 bp (base numbering according to published *M. trossulus* p53 tumour suppressor-like protein mRNA sequence, AY611471). Most (94%) of the late leukemic animals sampled in 2007 from cages in Burrard Inlet (British Columbia, Canada) had the same p53-like genotype, C182T G392G C821T, whereas 75% of the healthy animals were homozygous at positions C182C and T821T, independent of the combination at the 392 bp position. As well, an increased number of allele variants in the leukemic animals were detected that may arise from somatic mutation events in haemocyte precursors or from additional p53-like gene copies in polyploidy. In a subsequent survey done in 2009 on mussels sampled directly from beaches near Horseshoe Bay and Jericho (Vancouver area, BC) and at Hopkins Beach (Sunshine Coast, BC), the p53 SNPs at the three positions were all heterozygous regardless of health status for almost all the individuals analyzed. Therefore, the correlation of these SNPs with mussel leukemia was found only in the caged mussels.

2.2 Materials and methods

2.2.1 Animals

Mussels (1152 in total) 40-60 mm in length were collected from the intertidal zone during the low tide level in February 2007 from Hopkins Beach (Sunshine Coast, British Columbia, Canada) and placed into net socks, 12 mussels per sock. These socks were attached to plastic frames, 16 socks per frame, that were subsequently submerged between depths of 5 and 15 m

below the water surface at the monitoring station M in Burrard Inlet off the coast of West Vancouver and at the R station off Hopkins Beach (Figure 2.1). In October 2007, all mussels were removed from these frames, transported to the laboratory and stored in aerated seawater collected from the sampling location at the water temperature measured in the field (10-11^{o}C) for up to 3 days until analysis. In March of 2009, mussels were collected from beaches near Horseshoe Bay and Jericho (Vancouver area, BC) and at Hopkins Beach (Sunshine Coast, BC), 120 mussels from each site and exposed on submerged frames at the monitoring station M in Burrard Inlet from March to October. The exposure experiment and genotyping was done in the same manner as in 2007. All mussels were confirmed to be *M. trossulus* by examining their internal transcribed spacer region fragment electrophoresis patterns using methods described in (Heath et al. 1995) (data not shown).

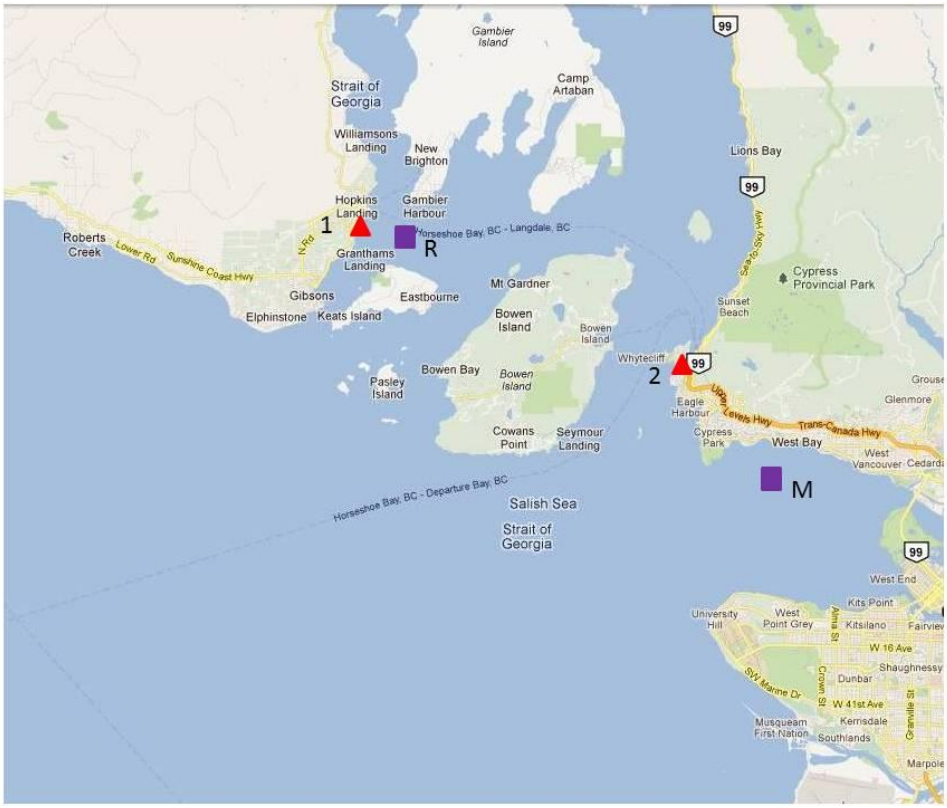


Figure 2.1. Mussel exposure experiment in Vancouver area (BC, Canada). ▲– collection sites, 1- Hopkins beach, remote site, 2- Horseshoe Bay, urban site. ■ – exposure sites, M – monitoring or urban site, R- reference or remote site.

2.2.2 Hemolymph sampling and assessment

Between 0.9 and 2.0 ml of hemolymph was withdrawn from the posterior adductor muscle into a 3 ml syringe preloaded with 0.5 ml of filtered, pre-chilled seawater. Haemocyte composition was assessed by examining a drop of hemolymph, which was placed on a glass slide pre-coated with poly-L-Lysine. Coated slides were prepared by washing the slides for 2 hours in wash buffer (100 mL 99% ethanol (CAS 64-17-5) and 70 mL 6.25 M sodium hydroxide (CAS 1310-73-2)) at room temperature with shaking at 70rpm. Slides were then rinsed in distilled water 5 times, air dried, and dipped in a 0.01% poly-L-Lysine (Sigma-Aldrich, p8920) and 0.1x phosphate buffered saline (PBS) (pH 7.4) solution for 1 hour. Slides were then rinsed in distilled water two times and air-dried. Cells were allowed to attach to the coated slide for 10 min before being analyzed by phase contrast microscopy (40x, Axiostar plus, Zeiss). Cells in 3 to 5 microscope fields were counted, depending on the cell density, in order to determine the ratio of cell types in each sample. Samples containing large aggregates of cells (>30 cells), where it was impossible to distinguish cell types or count the number of cells in the aggregates, were not included in the analysis. Cell types were categorized as described in Section 2.2.3. The remaining cells from the hemolymph samples were collected by centrifugation at 448 g for 6 min at 4^{o}C. The supernatant was discarded and the cell pellet kept frozen at -80^{o}C until analysis.

2.2.3 Cell types and phenotypic animal groups

Bivalve haemocytes are classified according to their morphology into granulocytes and agranulocytes (Dyrynda et al. 1997, Hine 1999, Pipe 1990, Pipe et al. 1997, Ratcliffe and Rowley 1981). For identification of these cell types visual appearance of stained (Wright-Giemsa stain, modified, Sigma-Aldrich, procedure No. WG, horizontal staining) and unstained cells under phase-contrast microscopy, 40x and 100x magnification was used. Although there are sub-groups of granulocytes shown to be immunogenetically different (Carballal et al. 1997, Dyrynda et al. 1997), all granulocytes were grouped together. Granulocytes cells that were large (7-12 µm) with conspicuous granules in their cytoplasm, a low nucleus to cytoplasm ratio and large spread pseudopodia when attached to the microscope slide (Barber 2004, Bower 1989, Ratcliffe and Rowley 1981) (indicated as G on Figure 2.2) were classified as normal. As well, several types of agranulocytes can be distinguished

(Carballal et al. 1997, Dyrynda et al. 1997, Hine 1999, Pipe 1990), and these were combined into another group based on their common attributes. Normal agranulocytes (A) were classified if they were generally smaller (4-6 µm) than granulocytes and did not have spread pseudopodia but much smaller protrusions (indicated as A in Figure 2.2a). Also, agranulocytes were characterized by an absence of granules and a low cytoplasm volume.

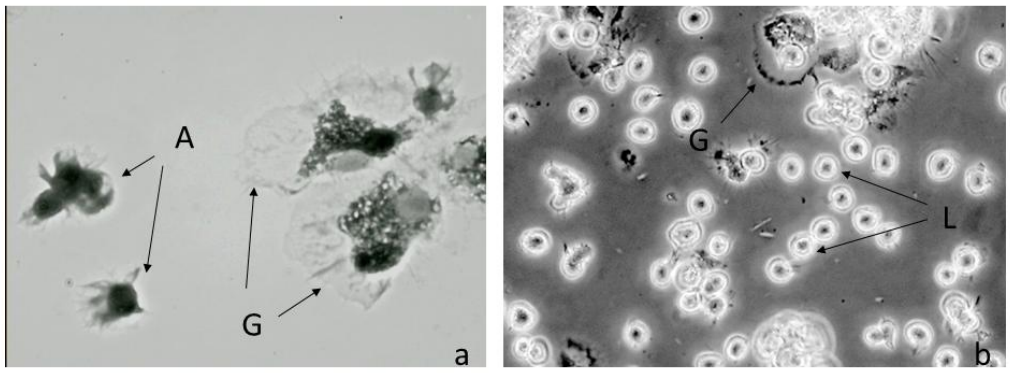


Figure 2.2. Haemocytological preparations of mussel blood cells showing the different cell types: granulocytes (G), agranulocytes (A) and neoplastic cells (L). Cells in (a) were stained with Wright-Giemsa stain. (phase contrast, 40x).

Neoplastic cells, in advanced stages of the disease, are easy to recognize since these cells are morphologically very distinct. Neoplastic cells are round and smooth due to their inability to adhere or spread and have very high nucleus to cytoplasm ratios (Barber 2004, Bower 2006, Elston et al. 1988, Smolowitz et al. 1989). Cells with these features were classified as leukemic (indicated as L in Figure 2.2 b). Late stage leukemia is easy to recognize due to the high density of neoplastic cells. However, leukemia is difficult to recognize at the early stages since some neoplastic cells share morphological similarities with normal cells, as has been demonstrated in other studies using antibodies developed against leukemic cells (McGladdery et al. 2001, Smolowitz et al. 1989). Thus, false negatives for early stages of the disease are possible, which makes visual haematocytology unsuitable for early diagnosis of leukemia.

For assessing the leukemia stage at least 100 cells were counted in at least 3 non-overlapping microscope fields. The ratio of neoplastic-looking cells to the total number of cells was used as a leukemia stage measure.

Samples were grouped according to the ratio of neoplastic cells to the total cells count. The normal phenotype included animals with hemolymph having less than 5% of leukemic-appearing cells. The late leukemic phenotype included animals with hemolymph with more than 90% of leukemic cells. The transitional phenotype included animals with hemolymph with anything from 5 to 90% leukemic cells and these animals were considered to be at intermediate stages of the disease.

2.2.4 Detection of single nucleotide polymorphisms

Total RNA was extracted from 106 hemolymph samples with E.Z.N.A. Total RNA Kit (Omega Bio- tech, USA, R6834-02). The cDNA was synthesized with Superscript II™ reverse transcriptase with oligo-dT primers according to the manufacturer's instructions (Invitrogen Life Technologies, Mississauga, ON). In order to detect single nucleotide polymorphisms (SNPs) the 4-1389 bp fragment of the mussel p53 mRNA coding region (base numbering according to published *M. trossulus* p53 tumour suppressor-like protein mRNA sequence, AY611471) was PCR amplified using the forward primer 4F and the reverse primer 1389R (Table 2.1, Figure 2.3). After locating the SNPs a smaller fragment 121-921bp was amplified with the forward primer F121 and the reverse primer R921 (Table 2.1, Figure 2.3) for genotype screening. This fragment was chosen since it is the shortest sequence containing all the SNPs that were found in the p53 and p63/p73 cDNA sequence. All PCR reactions were performed with iProof High Fidelity Polymerase (Bio-Rad, #172-5302, Hercules, CA) in the presence of 0.2 mM of $MgCl_2$ using the following protocol. Initial denaturation at 98^{o}C for 30 sec, followed by 33 cycles with denaturation at 98^{o}C for 10 sec, annealing for 20 sec at an initial temperature of 62^{o}C, which was decreased by 1^{o} C per cycle across 8 cycles followed by an annealing temperature of 54^{o}C for 25 cycles, and primer extension at 72^{o}C for 35 sec. After completion of all 33 cycles, final extension was at 72^{o}C for 7 min.

The amplified fragments were sequenced by the McGill University and Genome Quebec Innovation Centre Sequencing Facility (NANUQ, 740 Dr Penfield Avenue, Montréal (Québec) Canada). The sequencing was done in two directions for the initial SNPs identification and only in a single direction for the shorter fragment 121-921bp. The sequencing results were compared with the published *M. trossulus* p53 tumor suppressor-like protein cDNA sequence, AY611471, using Staden Package 1.5 and ClustalX 1.83. Heterozygotes were called where the height of the second peak was at least one fourth of the height of the main peak on the chromatogram.

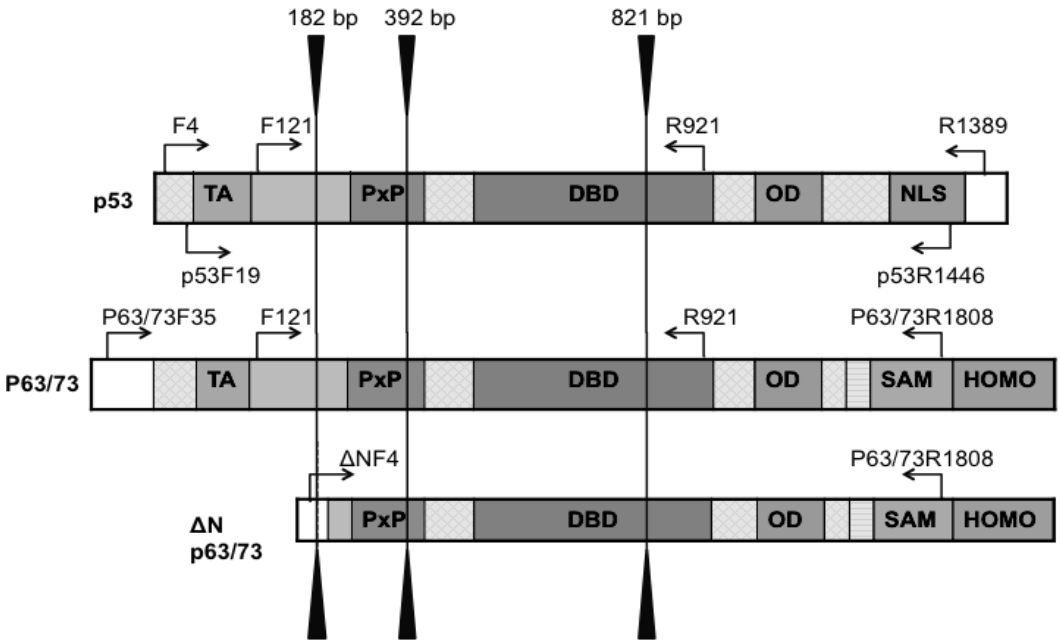


Figure 2.3. Locations of primer binding sites within the protein domains of p53, p63/73 and ΔNp63/73. TA – transactivation domain, PxP – proline-rich domain, DBD – DNA binding domain, OD – oligomerization domain. Bold vertical lines indicate the variable positions.

2.2.5 p53 isoforms analysis

Based on the published sequences of *M. trossulus* p53 isoforms, p53 (AY579471), p63/p73 (DQ865152) and ΔNp63/p73 (DQ865153), primers were developed that allowed for amplification of individual isoforms. The positions of these primers are shown in Figure 2.3 and sequences are presented in Table 2.1. Amplification of the correct PCR product was confirmed by observing a single band of the expected length on an electrophoresed agarose gel. Shorter PCR fragments were re-amplified from each individual isoform using primers F121 and R921 (see Section 2.4) for p53 and p63/p73 and F312 (Table 2.1) and R921(see Section 2.4) for ΔNp63/p73, ligated to a pJET1.2 vector (CloneJET™ PCR Cloning Kit, Fermentas, # K1232) and transformed into TOP10 cells (One Shot® TOP10 Chemically Competent *E. coli, Invitrogen,* C4040-03). Inserts from individual colonies were amplified using pJET forward and reverse primers and sequenced in two directions by the McGill University and Genome Quebec Innovation Centre Sequencing Facility (NANUQ).

Table 2.1. *Mytilus trossulus* p53 isoform-specific primers.

Primer name	*Primer sequence, 5'-3'.*
4F	TAATGTCACAAGCTTCAGTTTC
1389R	GTCCAAAAATGAATGCGCCAC
F121	CAAACATCACATCTAAAGAATCAATCG
R921	TCACTTCTACGGCACGTC
F312	GCATCAAGTCCTTACAATGATACA
p53F19	CAGTTTCAACTACATGCACACC
P53R1446	TGCTTTCAGGACGCATCT
p73F35	CAATATGGAACTCCCGTCCAGA
p73R1808	CTAATGGTGGATGCAGTGCT
ΔNp73 F4	ATCAAATTTGAGAGAACTGGATTTACAACC

2.2.6 Statistical analysis

Statistical analysis was used to test for associations between genotype and phenotype using the software *JMP*, Version 5.1.2 (Cary, NC, USA). It was assumed that the analyzed mussels came from a homogeneous population since they were collected from the same beach location. A standard chi-squared test under the null hypothesis of no association was applied:

$$\chi^2 = \frac{\sum_j [O_i - E_i]_j}{E_i} \quad (1)$$

where O_{ij} is the observed count for each phenotype (i) in each genotypic group (j), and E_{ij} is the expected count for each case in each group defined as follows:

$$E_{ij} = n_i m_j / N \quad (2)$$

where n and m are total counts for the case i and the animal group j respectively and N is a total amount of animals that were analyzed.

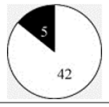
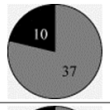
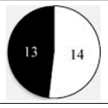
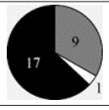
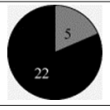
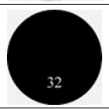
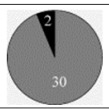
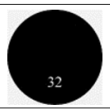
2.3 Results

2.3.1 p53-like cDNA sequence variations

Out of the 106 individual mussels caged in 2007 that were analyzed for leukemia status, 47 were classified as phenotypically normal, 27 as transitional and 32 as late-stage leukemic. An initial survey of a subset of 60 animals by sequencing of the F4-R1389 amplicons revealed three sites with SNPs at positions 182, 392 and 821 bp. All of the polymorphisms observed at these three positions were synonymous substitutions. Using the shorter F121-R921 amplicons from all 106 animals the distribution of SNPs among the phenotypic groups was surveyed (Table 2.2). At each variable position, the nucleotide present in the reference sequence is underlined in the reference sequence fragment given in parentheses below the location base number.

The 182 bp site is upstream of the proline-rich domain in a highly variable region (Muttray et al. 2005). Substitutions between C and T were detected at this position, which causes a change from an AGC (61Ser) codon with usage frequency of 20 per 1000 codons to an AGT (61Ser) codon with a lower usage frequency of 6.8 per 1000 codons. Out of mussels caged in 2007, all late leukemic animals, 48% of transitional and 11% of normal animals were heterozygous (C/T). The remaining animals in each of the three groups were found to be homozygous (T/T), and no C/C homozygote was detected at this position.

Table 2.2. Distribution of p53 cDNA sequence variations at each of the three polymorphic sites (182, 392 and 821 bp) in the three phenotypic groups; normal, transitional and leukemic in mussels caged in 2007.

Phenotype (number of animals)	Variable positions (underlined nucleotide in reference sequence)		
	182 (CCAG<u>C</u>ATAT)	392 (CCTC<u>T</u>GTGC)	821 (TGTG<u>T</u>CTGG)
Normal (47)			
Transitional (27)			
Leukemic (32)			
Legend	C/T T/T	G/G G/T T/T	C/C C/T

The 392 bp site is at the end of the highly conserved proline-rich domain. There, variations between T and G were detected, which change a TCT (129Ser) codon with usage frequency 6.8 per 1000 codons to a TCG (Ser) codon with the same usage frequency of 6.8 per 1000 codons. Guanine at this position creates CpG pairs (adjacent C and G in the 5' to 3' direction) on both DNA strands, which are potential sites for methylation. Out of mussels caged in 2007, most (94%) late leukemic animals were homozygous (G/G) at this locus, while 63% of transitional animals and 49% of normal animals were heterozygous (G/T).

The 821 bp site is located at the end of the highly conserved DNA binding region in the area involved with Zn-binding. Unlike the previous two SNPs this one is located in close proximity to a mutation hotspot of the human p53 gene, which is in the fourth DNA binding domain region (IARC TP53 database**,** R14, November 2009 (Petitjean et al. 2007)**).** Variation between T and C at this position changes a TGT (272 Cys) with usage frequency of 10 per 1000 codons to TGC (Cys) codon with much higher usage frequency of 95 per 1000 codons. Sequences with C at this location have an extra site for potential methylation on the coding DNA strand (CCTGG). Out of mussels caged in 2007 all leukemic, 81% of transitional and 21% of normal animals were heterozygous at this position. Homozygote TT was not detected at this position.

2.3.2 Association of cDNA variants with phenotype for mussels caged in 2007

Eight different combinations were found of the three SNPs at positions 182, 392 and 821 bp from the start codon (Table 2.3). 30 out of the 32 late leukemic animals had the same genotype of C182T G392G C821T. For these particular hemolymph samples, no normal cells in the drop were observed that was viewed under the microscope and therefore it was assumed that the vast majority of cells analyzed from these samples were neoplastic. The other two leukemic animals had a genotype that was different at the 392 bp position: C182T G392T C821T. Interestingly, for these two late-stage leukemic animals, some normal cells (< 5%) were also observed in the aliquot of hemolymph viewed under the microscope. Since among the mussels caged in 2007 only very few (7%) of normal phenotypes had the genotype C182T G392G C821T, it was hypothesized that this genotype is associated with the presence of neoplastic cells at least in the samples from the submerged frames. To test the strength of the association of leukemia with genotype C182T G392G C821T the chi-squared test with expected and observed counts was used as presented in Table 2.4. The statistical test showed that the association between the genotype C182T G392G C821T and late-stage leukemia is

statistically significant in caged mussels ($\chi^2 = 56.064$; $p < 0.01$) and from now on the genotype C182T G392T C821T will be referred as the leukemic genotype.

Table 2.3. Percentage of animals caged in 2007 in each phenotypic group (normal, transitional and leukemic) with the various genotypes given in the left hand side columns.

Genotype			Percentage of animals in each phenotypic group		
182	392	821	Normal	Transitional	Leukemic
T/T	G/T	C/C	38	16	0
T/T	T/T	C/C	23	16	0
T/T	G/G	C/C	14	5	0
T/T	G/G	C/C	2	0	0
T/T	G/T	C/T	13	21	0
C/T	G/T	C/T	4	26	6
C/T	G/G	C/T	6	16	94

A total of 8 genotypes were identified (Table 2.3) and all of these were present in the normal and transitional phenotypes. A trend was observed in the phenotypically normal animals where most (75%) had genotypes with homozygotes at the 182 bp and 821 bp positions. Using the chi-squared test the association between genotypes homozygous at the 182 and 821 bp positions and the normal phenotype was significant (χ^2=35, $p < 0.01$).

No clear trend emerged between genotype and the transitional phenotype, i.e. animals with a mixture of normal and leukemic cells in their hemolymph. However, since the leukemic phenotype was associated with heterozygotes at 182 and 831 bp, appearance of heterozygotes at one or both of these positions were observed in 63% of the transitional phenotypes. The transitional phenotype represents early stages of HN recognized from the appearance of neoplastic cells with normal cells still prevalent. The animals that were classified transitional in this study had a wide range of cell ratios in their hemolymph (from 5% to 95% leukemic cells). It was difficult, using visual inspection, to classify with great confidence as transitional, samples that had low percentages of neoplastic-appearing cells. Nevertheless, the strength of association between genotypes with heterozygosity at 821 bp and presence of neoplastic cells was tested, i.e. for both leukemic and transitional phenotypes (Table 2.4) and a statistically significant association was found.

Table 2.4. Case-control counts for testing the statistical significance of associations between phenotypic and genotypic groups for 2007 caged samples. Numbers represent number of animals.

Phenotype	**Leukemic genotype** (T182C G392G C821T)	
	Present	Absent
Leukemic	30	2
Normal	2	47
Chi squared = 61, P < 0.01		
Normal phenotype	**Normal genotypes** (T182T (G392G, G392T, T392T) C821C)	
	Present	Absent
Present	37	12
Absent	10	49
Chi squared = 35, p < 0.001		
Presence of neoplastic cells (leukemic or transitional phenotype)	**Early stage leukemia genotypes** ((T182T, C182T) (G392T, G392G) C821T)	
	Present	Absent
Present	49	10
Absent	12	35
Chi-squared = 33, p < 0.01		

2.3.3 Haplotypes for each of the known p53 isoforms

In the preliminary screening for SNPs, sequencing was performed directly on PCR amplification products, which include both full-length isoforms, p53 and p63/p73. This approach does not yield information on the number of haplotypes nor does it tell us to which p53-isoform the variations are attributable. Thus, in an additional analysis, isoform specific sequences were cloned in *Escherichia coli* and then several colonies from each sample and isoform were sequenced. This was performed on a selection of four leukemic and four normal samples using the original target cDNA and primers specific for the three p53-like isoforms, p53 (AY579471), p63/p73 (DQ865152) and ΔNp63/p73 (DQ865153), as described in the Materials and Methods section. Since this analysis required 2 PCR steps, it is acknowledged that reaction conditions could have biased amplification of some products over others. The complete results from sequencing of a total of 216 clones are presented in Table 2.5. Two haplotypes were most frequent in the normal phenotype (109 out of 111 clones sequenced from the normal phenotypes). Two other alleles were found, 182C 392T 821T and 182C 392T 821C, however only in one clone each. The haplotypes agreed with the genotypes found from direct sequencing for 3 out of 4 normal phenotypes. For one normal animal the DNA trace from direct sequencing had only one peak for thymine at 392 bp although the haplotype 182T

392G 821C was found in 3 out of a total of 22 clones sequenced for this animal. The number of most frequently detected haplotypes increased to 4 for the leukemic phenotype. All of the haplotypes agreed with the leukemic genotype, C182T G392G C821T. One rare haplotype, 182C 392T 821C, was also detected in one of the leukemic samples. The allele variant 182T 392G 821C was common to both normal and leukemic phenotypes. However, the haplotypes 182C 392G 821T, 182T 392G 821T and 182C 392G 821C, were unique to the leukemic phenotype. The other haplotype, 182T 392T 831C, found in the normal phenotype was not found in any leukemic animals.

Table 2.5 cDNA haplotypes (rows) at the three polymorphic sites (182 bp, 392 bp and 821 bp, columns) in p53-like isoforms (columns) in leukemic (L) and normal (N) animals. Numbers are amounts of clones detected for the isoform in the animal.

Genotype (direct sequencing)			Haplotypes (cloned)					
			SNPs positions			p53 isoforms		
182	392	821	182	392	821	P53	P63/73	ΔNp73
Animal 1, N								
T/T	G/T	C/C	T	G	C	9	5	
			T	T	C	11		
				T	C			1
Animal 2, N								
T/T	G/G	C/C	T	G	C	2	6	
				T	C			9
Animal 3, N								
T/T	T/T	C/C	T	G	C	3		
			T	T	C	5	13	
			C	T	T	1		
				T	C			4
				G	C			1
				T	T			1
Animal 4, N								
T/T	G/T	C/C	T	G	C	6	5	
			T	T	C	4		
			C	T	C	1		
				T	C			13
				G	C			1

Table 2.5, continued. cDNA haplotypes (rows) at the three polymorphic sites (182 bp, 392 bp and 821 bp, columns) in p53-like isoforms (columns) in leukemic (L) and normal (N) animals. Numbers are amounts of clones detected for the isoform in the animal.

Genotype (direct sequencing)			Haplotypes (cloned)					
			SNPs positions			p53 isoforms		
Animal 5, L								
C/T	G/G	C/T	T	G	C	2	1	
			C	G	T	2	1	
			T	G	T	1		
			C	G	C	1	2	
			C	T	C		1	
				T	C			17
				G	C			6
				T	T			1
Animal 6, L								
C/T	G/G	C/T	T	G	C		2	
			C	G	T	1		
			T	G	T	1	2	
			C	G	C		2	
				T	C			5
				G	C			1
				T	T			2
Animal 7, L								
C/T	G/G	C/T	T	G	C	1		
			C	G	T	2	1	
			T	G	T		2	
			C	G	C	2	1	
				T	C			3
				G	C			2
Animal 8,L								
C/T	G/G	C/T	T	G	C	2	12	
			C	G	T	5		
			T	G	T	2		
			C	G	C	3		
				T	C			15
				G	C			1

Both normal haplotypes were found in the p53 isoform specific clones, but the p63/p73 isoform had only one haplotype in each of the normal animals. For 3 out of 4 normal animals, the p63/p73 haplotype was 182T 392G 821C, however in one normal it was 182T

392T 821C instead. In contrast, both p53 and p63/p73 had multiple haplotypes in leukemic animals. Therefore, it appears that the full-length p63/p73 isoform in normal animals is coded by a single allele in contrast to p53, which is coded by several alleles with greater sequence variability. With regards to ΔNp63/73, the majority of clones (27 out of 30) had the same haplotype (392T 921C) with a total of 3 different haplotypes observed. Interestingly, in all but one of the normal animals, the allele for the N terminal truncated isoform ΔNp63/p73 is different than that for its full-length version p63/p73 (182T 392G 821C).

2.3.4 Genotyping of *M. trossulus* intertidal populations sampled directly from the beach in 2009.

To explore if similar associations between p53 amplicon sequence variations and disease could be found in populations collected directly from the wild intertidal populations samples were analyzed that came directly from Hopkins Beach, Jericho and Horseshoe Bay (see Methods). More samples were collected in March 2009 and exposed on frames until October 2009 exactly in the same manner as was done in 2007. All these samples (155 in total) were genotyped as described in the Methods.

Table 2.6. Genotypes of 106 animals from beaches and from frames sampled in 2009. Percentage of animals in each phenotypic group (normal, transitional and leukemic) with the various genotypes given in the left hand side columns.

	Genotype			Percentage of animals in each phenotypic group		
	182	392	821	Normal	Transitional	Leukemic
beach	TT	GT	CT	2	0	0
	CT	GT	CT	7	6	0
	CT	GG	CT	91	94	100
frames	TT	TT, GT, GG	CC	55	0	0
	TT	TT, GT, GG	CT	15	67	0
	CT	GT	CT	25	33	0
	CT	GG	CT	5	0	100

The results of genotyping are presented in Table 2.6, where data for all beach locations is pooled together. It was found that the genotype with homozygotes at both flanking positions 182 and 821 bp or the “normal genotype” (T182T (G392G, G392T, T392T) C821C) was not detected in intertidal mussel population and only one animal had genotype T182T G392T C821T. All other intertidal animals were heterozygous at both flanking positions 182 and 821 with combinations at the middle position G392G and G392T. Both late transitional (>85% of neoplasic cells) and late leukemic groups had “leukemic” genotype C182T G392G C821T. However many normal (according to microscopy) animals had this “leukemic” genotype also.

Therefore, it was found that the same association of SNPs in p53 coding region and leukemia was valid only for the population exposed on frames, but not for the population, collected from the beach. Similar trends as in 2007 were seen with respect to association of homozygous flanking locations with normal mussels in the 2009 caged samples, but the association was weaker in the 2009 samples. For both beach and caged samples the strong association between late-stage leukemia and 182CT 392GG 821CT was upheld.

2.4 Discussion

2.4.1 Association of the leukemic genotype C182T G392G C821T with leukemia

Three highly variable sites were discovered in p53-like cDNA sequences from *M. trossulus* normal and leukemic haemocytes. One particular combination of SNPs, namely C182T G392G C821T, was strongly correlated with late-stage leukemia (>95% of L cells, 30 samples) in animals sampled from the submerged frames. Since these are synonymous substitutions, it is unknown whether these SNPs lead to alterations of p53 pathways and deregulation of the tumor suppression function allowing leukemic cells to survive or whether this genotype-leukemia association is an indirect result of other processes involved in tumour proliferation. There are two factors supporting a causative role in leukemia for the detected variations in the p53 gene. First, p53 is a key suppressor of tumour progression and is dysfunctional in many cancers (Giaccia and Kastan 2006). Specifically, differential expression of the p53-family at both the mRNA and protein levels has been shown to be associated with bivalve leukemia (Kelley et al. 2001, Muttray et al. 2008{Kelley, 2001 #21)). As second, no homozygotes C182C and T821T were detected in any of the leukemic animals.

In contrast, most of the animals collected from the intertidal beach areas had the genotype C182T G392G C821T regardless of their disease status. Homozygotes at two the flanking positions, 182 and 821 bp, were detected in less than 10% of beach samples. Inter-tidal and submerged mussels are exposed to different environments and experience different stressors. In particular, intertidal emersion is known to be associated with desiccation, UV exposure and extreme temperatures. It was demonstrated that intertidal mussels can undergo significant physiological and biochemical changes, such as upregulation of stress-response genes (Snyder and Rossi 2004) or switching to anaerobic metabolism (Stillman and Somero 1996) during the emersion. The different environment and variety of stressors may lead to differential expression of p53 haplotypes. As well it was demonstrated that DNA damaging

stress such as UV exposure elevates the p53 expression and induces apoptosis, thus reversing the mussel leukemia (Bottger et al. 2008). It was hypothesized that the emersion stress might affect the phenotype-genotype association via reversing the disease and it might affect the differential expression of p53 alleles, leading to the haplotype loss in the intertidal animals. Therefore, the usefulness of genetic biomarkers can vary greatly depending on the environment and location.

2.4.2 Haplotypes in p53 isoforms

The current state of knowledge suggests that there is only one p53-like gene copy present in most invertebrates, including bivalves (Belyi et al. 2010, Lu et al. 2009, Stifanic et al. 2009). Sequence analysis of the first invertebrate p53-like isoforms, which were found in *M. arenaria*, allowed Kelly and co-authors (Nedelcu and Tan 2007) to suggest that a single gene was capable of producing the two proteins, one p53-like and the other p63/p73-like, both of which were differentially expressed in normal and leukemic haemocytes. This was confirmed by an analysis of the p53 gene structure in *M. arenaria* (FJ041332, (Kelley et al. 2008)). Both the *M. arenaria* p53 mRNA (AF253323) and p73 mRNA (AF253324) sequences (Kelley et al. 2001) align with the genomic sequence (FJ041332) according to an alignment done with Spidey (Wheelan et al. 2001), which shows that this single gene can code for all isoforms. *M. arenaria* and *M. trossulus* p53 isoforms share a high degree of sequence similarity at the mRNA level and are phylogenetically closely related (Muttray et al. 2005). Given this, the multitude of p53-like haplotypes found in *M. trossulus* with HN was unexpected (Table 2.4). It was hypothesized that the high variability observed in the haplotypes might stem from somatic mutations. This may be the result of independent mutations in haemocyte precursors, which then give rise to multiple clonal populations of haemocytes. This process has been described for mammalian cancers (Tiu et al. 2007). Alternatively, the detected variations might be acquired by mature neoplastic cells along with an increasing level of ploidy and, as a consequence, an increasing number of p53 gene copies (Moore 1993). Another unexpected result was that the truncated ΔNp63/73 isoform is not always coded by the same allele as the full-length p63/73 isoform. In mammals, the truncated ΔNp63/73 isoform is known to be expressed from the p73 gene (Benard et al. 2003).

2.4.3 Possible mechanisms for effects of SNPs on p53 function

All of the SNPs that were detected are silent substitutions. Thus, the potential effect of these SNPs on p53 activity and possibly on disease development would be realized not at the protein level, but earlier, at the transcriptional or RNA processing stages. Here, it was suggested how these SNPs might affect expression of p53 family members.

Variations in the p53-like coding region sequence might cause changes in the secondary structure of the mRNA molecule, affecting its interaction with cellular components for RNA processing, transport, stability and expression (Shabalina et al. 2006). Any of these events may change the amount of p53 expressed in the cell (Babendure et al. 2009, Latchman 2002), causing the disruption of the regulatory pathways and allowing tumorigenesis to proceed.

SNPs are also known to induce alterations in the splicing process, affecting the structure and variants of the resulting proteins (Chao et al. 2001). It was shown in mammalian cells that the balance between p53 family isoforms is important for a functionally correct response to cellular stress or tumour suppression (Bourdon et al. 2005). Thus, alteration of splicing processes may affect p53 family activity in such a way that promotes tumor development. All these events can counteract the effect of transcriptional up-regulation and, as a result, lead to insufficient protein levels and the possibility of tumorigenesis.

DNA methylation is known to be one of the key mechanisms of gene expression regulation. Sequence variations at methylation sites are frequent for the p53 gene in human cancers (Greenblatt et al. 1994). In mussels, substitutions from T to G at the position 392 bp and from T to C at position the 821 bp create additional methylation sites, CpG on both DNA strands at the 392 bp position and a CCTGG site at the 821 bp position. The presence of additional methylation sites could significantly affect p53 expression if these sites become methylated.

Another possible effect of the detected SNPs is the alteration of the codon usage frequency, which in turn, changes the speed of translation and, as a consequence, might alter the protein conformation (Cortazzo et al. 2002). The presence of C at the 182 bp position and C at the 821 bp position create codons with much higher usage frequency in mussels. Position 821 is located in the highly conserved DNA binding region, therefore even small variations in the speed of protein translation and hence of protein conformation might have a significant

effect on the p53 protein activity. Effect of the codon frequency change at position 182 bp is not clear since it is located in the area known to be less conserved.

2.4.4 p53 cDNA variation as a potential molecular-level tool for mussel leukemia detection in caged mussel populations

Based on the correlation between p53 cDNA variants with heterozygotes at positions 821bp and/or 182 bp and the presence of neoplastic cells in hemolymph of caged mussels, it was proposed that screening for SNPs in mussel p53-like cDNA would be a useful genetic tool for detecting leukemia for the mussels from the submerged frames. Analysis of haplotypes, which involves cloning and more sequencing steps, would not be feasible, economically nor time-wise, in a practical environmental monitoring setting. However, high throughput SNP screening of F121-R921 fragments directly from PCR amplified cDNA samples would be doable in mussel health monitoring programs. One of the possible methods will be the High Resolution Melting Analysis of amplicons presented in Appendix B. Although a strong association between the leukemic genotype and presence of late-stage leukemia was found in the caged mussel populations, the leukemic genotype would not be a useful biomarker for presence of the disease. This is because the dynamics of the disease are still unknown and because of the high mortality associated with late stage HN. Instead, the presence of homozygotes (T182T and C821C) could be considered as an indicator of absence of the disease. To test for the leukemia prevalence in population exposed on the submerged frames, the ratio of animals with heterozygote C821T or both heterozygotes C182T C821T to those with homozygotes T182T and C821C can be used.

The fact that association of the genotype variations with the leukemia were detected only in caged populations creates the significant limitation of the applicability of the proposed marker. It can be used only for the populations, exposed on the submerged frames and is not applicable for the monitoring of the beach populations.

3 P53 sequence polymorphisms in *Mytilus edulis*

3.1 Scope

Genetic biomarkers based on DNA sequence polymorphisms that are associated with diseases have the potential to improve sub-lethal endpoint detection in environmental monitoring programs. However, genetic associations may differ depending on the species of organism used. Two *Mytilus* species, *M. edulis* and *M. trossulus*, commonly used in environmental monitoring, differ in their susceptibility to a haemic neoplasia with *M. trossulus* being more susceptible. In this study, p53 sequence variations in normal and diseased *M. edulis* were compared with those previously reported for *M. trossulus*. Forty single nucleotide polymorphisms (SNPs) were detected in *M. edulis* as opposed to only three that were found in the *M. trossulus* p53 coding region. Many of the *M. edulis* p53 SNPs were associated with the presence of neoplasia. The disease-associated SNPs in *M. edulis* p53 formed blocks of high linkage that aligned more closely with *M. trossulus* than with normal *M. edulis*. Together this suggested the possibility of hybridization between *M. edulis* and *M. trossulus.* Phylogenetic analysis showed that p53 haplotypes predicted for *M. edulis* with late stage neoplasia were similar with p53 haplotypes for *M. trossulus*. It was proposed that hybridization followed by recombination leads to the partial introgression, which cannot be detected by the currently available RFLP markers. According to these findings, it was proposed that hybridization with *M. trossulus* was connected to the increased disease susceptibility of *M. edulis*. This fact should be taken into account when using *M. edulis* for environmental effects monitoring and farming.

3.2 Material and methods

3.2.1 Hemolymph collection and analysis

Mytilus edulis mussels larger than 45mm in length were purchased from the Island Scallops farm (Vancouver Island, BC, Canada) in February 2010, placed in net socks that were attached to plastic rectangular frames and submerged at monitoring stations in Burrard Inlet as was described in the previous chapter. After eight months of exposure, the mussels were removed from the frames and hemolymph samples collected from the posterior adductor muscle in order to assess leukemia status by light microscopy and collect cells for DNA

analysis. Processing of the samples and cell composition analysis were performed as described before (Vassilenko et al. 2010, see Chapter 2). Briefly, samples were classified as normal if they contained less than 10% round and non-adhesive cells, late stage leukemic if there were more than 90% round, non-adhesive neoplastic cells in high density and transitional if they had 10-90% of neoplastic-looking cells. Total RNA was extracted and cDNA was synthesized with oligo-dT primers. Amplification of cDNA was performed using primers designed for *M. edulis* p53 mRNA sequence based on the published sequence (GeneBank accession number *AY579472).* Sequence of the forward primer F111 was 5'GGGGGATACACAAACATCACATCTA3' and the reverse primer, MeR988, was 5'TGTTTGCATGGTGGGAGAGC3'. Conditions for the PCR amplification were the same as described previously (Vassilenko et al. 2010). Sanger sequencing was performed at the McGill University and Genome Quebec Innovation Centre Sequencing Facility (NANUQ), Montréal, Canada with the same primers as were used for the PCR amplification. The cDNA PCR amplification and amplicon sequencing were done in duplicate.

3.2.2 Species identification

Species identification was done with four genetic markers. They were the internal transcribed spacer (ITS) region between the 18S and 28S nuclear-r-DNA coding regions, (Heath et al. 1995), Glu-5' (gene encoding for the mussel polyphenolic adhesive protein) (Rawson et al. 1996), Mal-I (*Mytilus* anonymous locus one) (Rawson et al. 2001) and the mitochondrial large subunit (16S) ribosomal DNA (16S-mtDNA) (Rawson and Hilbish 1995). **Error! Reference source not found.** lists the primers and restrictases used for the analysis.

Table 3.1. List of primers used for *Mytilus* species identification.

Marker	*Primer name*	*Primer sequence, 5'-3'*	*Restrictase*
ITS1	F	GTTTCCGTAGGTGAACCTG	HhaI
	R	CTCGTCTGATCTGAGGTCG	
Glu 5	*JH-5 F*	GTAGGAACAAAGCATGAACCA	
	JH54R	GGGGGGATAAGTTTTCTTAGG	
Mal-I	F	GAAGCGTATTTGGTCACTGGCAC	SpeI
	R	GTCATAAAATGGAACATCTGAGTC	
16S	AR F	CGCCTGTTTATCAAAAACAT	SpeI,HaeIII,EcoRV
	BR R	CCGGTCTGAACTCAGATCACGT	

3.2.3 SNPs detection and analysis

Sequence assembly and SNP detection were done using the Staden Package1.5 (Pregap4, Gap4, http://staden.sourceforge.net/). Genotype analysis (linkage disequilibrium, haplotype prediction and association analysis was performed with Haploview (Barrett et al. 2005). Haplotype network was generated using the median-joining network algorithm using the software program Network version 4.600 (Bandelt et al. 1999). A phylogenetic tree showing the molecular evolutionary relationships between the predicted haplotypes for *M. trossulus* and *M. edulis* was constructed using the Neighbor-Joining method with the Jukes-Cantor model and 1,000 bootstraps.

3.3 Results and discussion

3.3.1 Species identity DNA markers

Currently, there are several PCR-based markers that are commonly used to distinguish between *M. edulis, M. trossulus* and their hybrids (Heath et al. 1995, Rawson et al. 2001). All samples used in this study were analyzed using the ITS marker and a random subgroup of 47 *M. edulis* and 20 *M. trossulus* were analyzed with all four markers, as described in Methods section. Representative data are shown in Figure 3.1. The ITS and 16S-mtDNA markers revealed the expected fragment patterns for each species, including the leukemic *M. edulis*. The Mal-I marker revealed two genotypes in *M. trossulus* and one genotype in *M. edulis.* No hybrid genotypes with mixed patterns, were detected by the ITS, 16S-mtDNA and Mal-I markers for any of the samples. According to the Glu 5' marker, *M. trossulus* had one genotype, while *M. edulis* had at least four genotypes. As well, the Glu 5' marker showed the presence of a *M. trossulus*-specific (250bp) fragment in 5 of the *M. edulis* samples in addition to the expected *M. edulis*-specific fragments. However, none of these five *M. edulis* samples had leukemia or the *trossulus*-like p53 SNPs. The Glu 5' marker was the only one that detected *M. trossulus* bands in the M. *edulis* patterns. It was concluded that *M. edulis* samples originate from populations containing hybrids with *M. trossulus*.

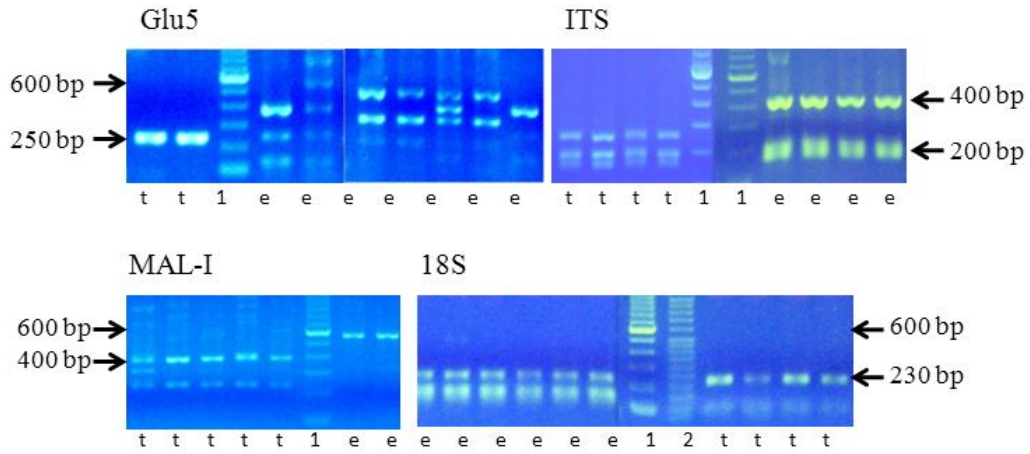


Figure 3.1. Mytilus species identification with 4 genetic markers, Glu5, ITS, Mal-I and18S. t-trossulus, e-edulis. DNA ladders (Invitrogen): 1- 100bp, 2-50bp.

3.3.2 SNPs detected in coding region of *M. edulis* p53 mRNA

A PCR amplified section of the p53 coding region was screened for SNPs in a total of 57 animals. Of these, 26 were classified as normal (i.e., their hemolymph contained less than 10% of neoplastic-looking cells), and the other 31 as diseased, including 9 late leukemic animals (i.e., with more than 90% of neoplastic-looking cells) and 21 transitional samples (i.e., with 10-90% of neoplastic-looking cells). In the amplified fragment, 40 variable positions were discovered, all of which were single nucleotide substitutions or polymorphisms (SNPs). The genotype data for each sample is shown in Appendix D. Frequencies of the dominant nucleotides at each SNP location and for the three phenotypic groups are shown in Figure 3.2. Only SNPs with minor allele frequencies of greater than 5%, 36 in total, are listed on the x axis. Reference nucleotides are according to the *M. edulis* p53 sequence published at GenBank, accession number AY579472. Out of the 40 SNPs, 37 were synonymous substitutions, while the other 3, located at 219, 261 and 346 bp, were non-synonymous. Several trends can be noticed on the chart. At the majority of SNP positions, the reference allele is very frequent (80-100%) in the normal and transitional groups, however quite rare (15-25%) in the late leukemic group. This indicates that the late leukemia stage is associated with the presence of a rare allele, while this is not necessarily so at earlier stages of the disease. The association of these SNPs with leukemia is discussed later.

The extraordinarily high number of variations in such a short fragment of what is considered to be a very conserved gene was surprising and cannot be explained by mutation processes alone. Sample sequences were compared with sequences published at GenBank for *M. trossulus* p53 mRNA (accession number AY611471) and *M. edulis* p53 mRNA (accession number AY579472). The p53 sequences of these two species are nearly identical and differ at a few single nucleotides sites. Alignment of the p53 mRNA sequences: those published in GenBank and those obtained from our samples, revealed that the most frequent *M. edulis* SNPs, which were 28 of a total of 40, exactly matched the positions where the GenBank AY611471 *M. trossulus* sequence differs from the GenBank AY579472 *M. edulis* sequence, which suggests a strong possibility of hybridization (Figure 3.3).

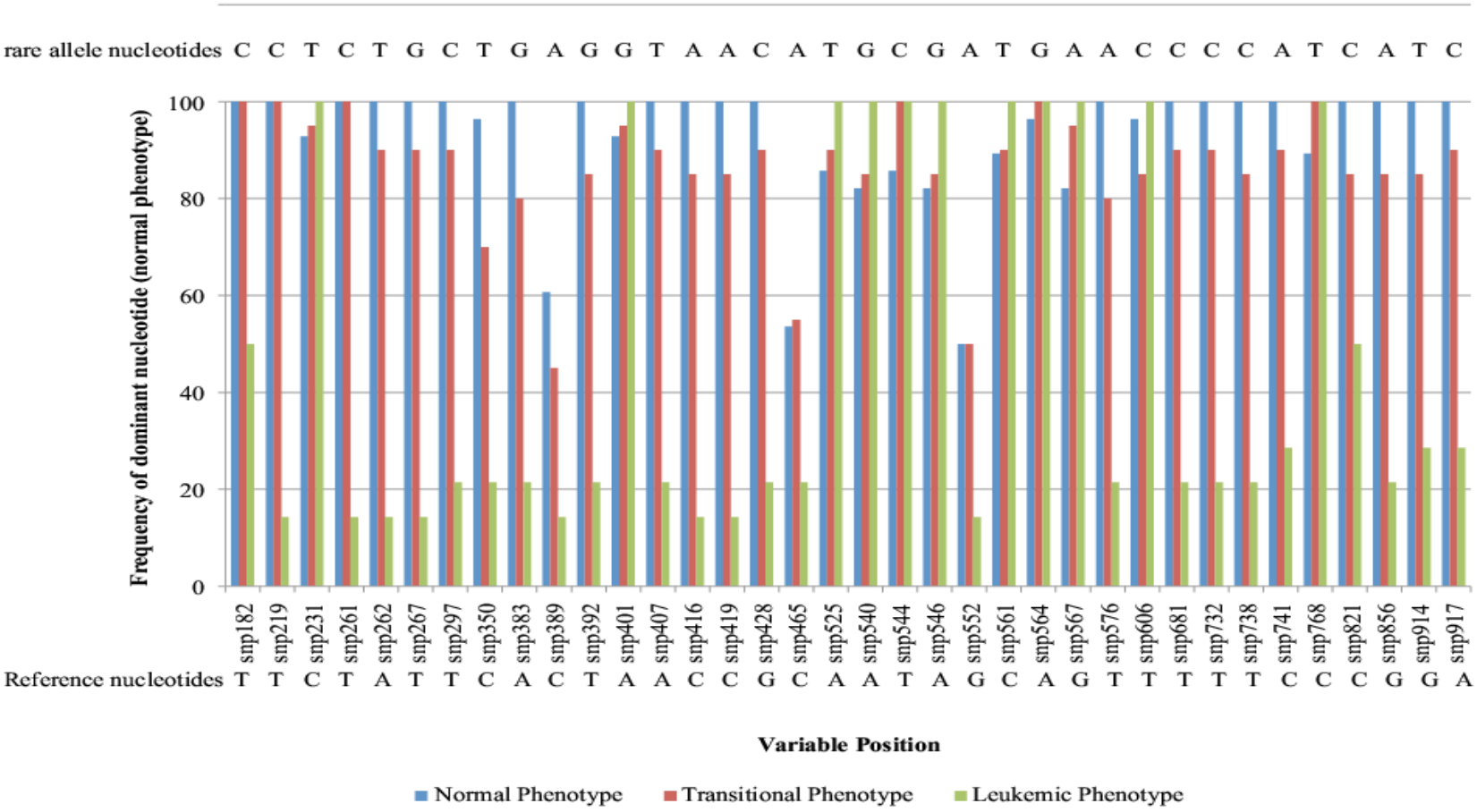


Figure 3.2. Genotype frequencies in the three phenotypic groups, normal(<10% of neoplastic looking cells), transitional (10-90% of neoplastic cells) and late leukemic (>90% of neoplastic cells) at each of the variable positions with MAF>10% detected at coding region of M. edulis p53 mRNA.

The remaining 12 SNPs were less frequent than the first 28 SNPs and did not present any pattern. Figure 3.3 shows a representative fragment of the genotype alignment of 6 late leukemic, 8 normal samples as well as the reference sequences for *M. trossulus* and *M. edulis*. The fragment includes five variable positions (shaded), four of which match the positions where *M. trossulus* and *M. edulis* are different.

Figure 3.3. Representative fragment from the genotype alignment of 6 late leukemic samples (edul1-6L), 8 non-diseased samples (edul7-14) and published *M. edulis* p53 mRNA (accession number AY579472)and *M. trossulus* p53 mRNA (accession number AY611471) sequences. Four out of five variable positions (shaded) exactly match the positions where published *M. trossulus* sequence differs from the published *M. edulis* sequence.

To further explore the hypothesis of hybridization a linkage disequilibrium (LD) analysis using Haploview 4.2 (Figure 3.4c) was performed.

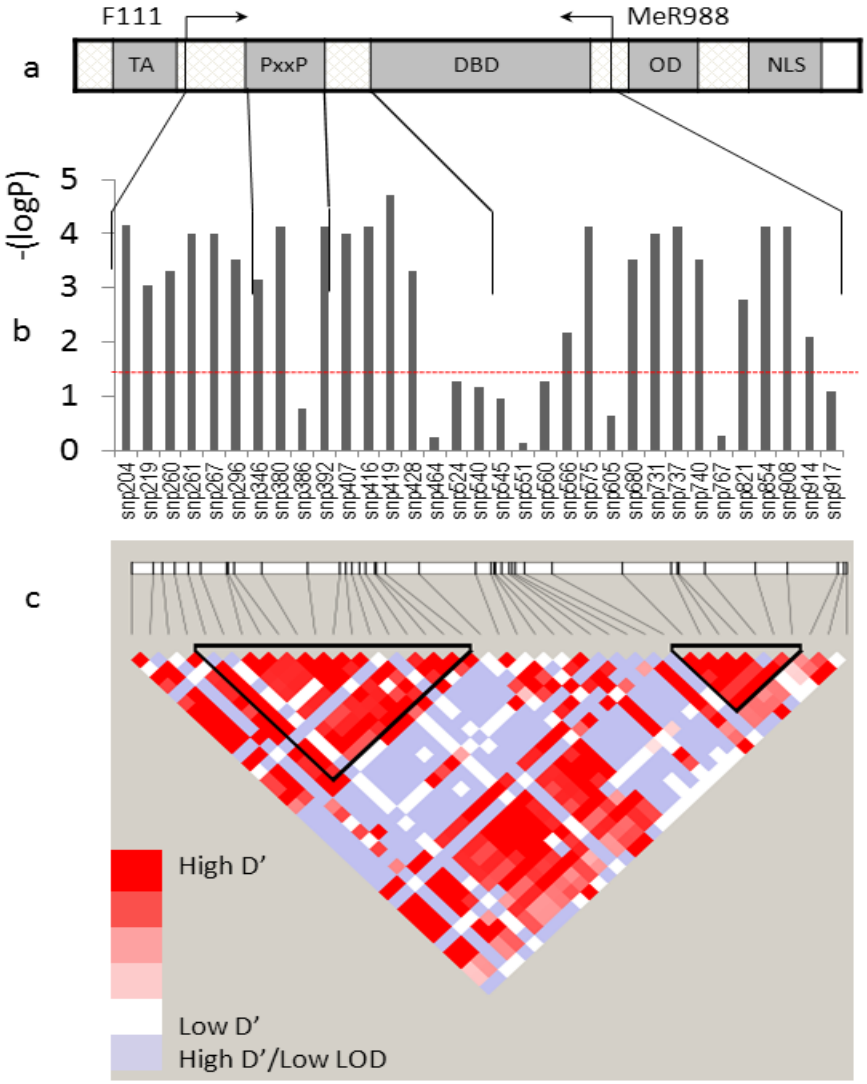


Figure 3.4. P53 variation sites detected in *M. edulis* cDNA coding region; a: p53 functional blocks : TA-transactivation domain, PXXP-proline-rich domain, DBD-DNA binding domain, OD-oligomerization domain, NLS –nuclear localization domain; b: association of the detected SNPs with late stage of HN; control group had non-diseased samples; red dotted line shows the association significance threshold (p=0.05); c: linkage disequilibrium plot (Haploview4.2) for the detected SNPs.

LD analysis revealed two blocks of SNPs with high linkage (depicted in red within the solid triangles), which are separated by an area of low linkage (depicted in grey). These two flanking blocks of high linkage are 260 and 174 bp long, respectively, and interestingly, they are in high linkage with each other. The area of low linkage that separates the two flanking

blocks, is 236 bp long. This pattern, where high linkage blocks are interrupted by a low linkage area, is similar to the *mosaic* pattern described for human genes (Daly et al. 2001, The_International_HapMap_Consortium 2003). High linkage blocks interrupted by a low linkage area suggest the presence of a recombination hotspot. Despite the fact that mosaic phenomena are known to occur in many genes, the presence of a recombination hotspot in the middle of the p53 coding region is surprising. This is remarkable as this region is thought to be highly conserved among biological taxa (Walker et al. 2011). Moreover, recombination within the p53 gene sequence is even more surprising since the gene in *Mytilus* sp. is expected to be relatively short (about 20kb) and according to Carlson et al. (Carlson et al. 2004) such short fragments should not undergo recombination. Even though the p53 genomic sequences remains unknown for *Mytilus sp.*, since it is a highly conserved gene, the *Mytilus* sp. p53 genes can be assumed to be of a length similar to p53 gene sizes for other species. For example, the genomic p53 gene for another bivalve species, *Mya arenaria* (accession number FJ041332.1), is 17kb and for *Homo sapiens* it is 26kb (as. num. NG_017013); therefore our estimate of 20kb is plausible.

3.3.3 Association with Haemic Neoplasia

Case-control association analyses of SNPs with leukemia in *M. edulis* were performed with two groups of animals, where the case group was late leukemic (9 samples) and the control group was non-diseased (23 samples). Associations of the individual SNPs with the disease are shown in Figure 3.4b, where each bar shows the –log(*p-value*) for the chi^2 test of independence of leukemia from the presence of rare allele. The significance threshold, p=0.05, is indicated by the dashed horizontal line. One can see that the SNPs that are in significant association with leukemia form two blocks, SNPs 7-19 (i.e. snp260, snp261, snp267, snp296, snp346, snp380, snp386, snp392, snp401, snp407, snp416, snp419) and SNPs 31-39 (i.e. snp680, snp731, snp737, snp740, snp767, snp821, snp854, snp908, snp914). SNPs in between these two blocks were not significantly associated with leukemia. Moreover, the flanking blocks of SNPs in significant association with the disease overlap with the high linkage blocks, while the middle block of SNPs with no association overlaps with the low linkage area (Figure 3.4c). Including the transitional samples to the case group decreases the association of the rare allele with the disease, as will be shown below.

3.3.4 TagSNPs

To examine whether some of these SNPs could be used for a *M. edulis* specific leukemia biomarker, the set of SNPs was narrowed to include only those with a strong association with the disease. For the two phenotypic groups, non-diseased and late leukemic, a set of SNPs with the best association with the disease was selected. They were SNP261 (p= 1.49E-12) and SNP419 (p = 1.49E-12). Haplotype 261A 419C was detected in all samples with normal phenotype and 16.7% of late leukemic samples (Table 3.2). Haplotype 261T 419A was detected in the remaining of 83.3% of late leukemic samples (Table 3.2, left side). However, to increase the sensitivity of the genetic biomarker, it would have to detect early stages of the disease. Therefore, the transitional group was included into the analysis, i.e. all samples presenting more than 10% neoplastic cells were used for the diseased, or affected group, which significantly changed the association. Two more haplotypes were discovered, 261T 419C and 261A 419A, however they had low frequencies, even though the former contains both reference alleles (Figure 3.2). In the new analysis, haplotype 261A 419C was detected not only in all non-diseased samples, but also in 62.4% of diseased samples. The three haplotypes 261T 419A, 261T 419C and 261A 419A (Table 3.2), which were strongly associated with late-stage leukemia were found in only 37.6% of all diseased samples (late-stage and transitional samples), which makes these haplotypes useful only for detecting late-stages. Lack of any significant associations of SNPs with all stages of the disease might have arisen from inaccuracies in the detection of early leukemia stages using cell morphology and/or from the fact that genetic variations or expression of alleles changes as the disease progresses. Inaccuracies in leukemia detection, in particular false-positive detection, may have occurred due to the miss-interpretation of agranulocytes with few pseudopodia for the neoplastic cells.

Table 3.2. Significance of association of tagSNPs with mussel leukemia and case-control frequencies.

	Non-diseased group versus late leukemic					Non-diseased group versus all leukemia stages				
snp261 snp419	Total Haplotype frequencies	Haplotype freq. in affected group	Haplotype freq. in non-affected group	Chi ^{2}	P value	Total Haplotype frequencies	Haplotype freq. in affected group	Haplotype freq. in non-affected group	Chi ^{2}	P value
AC	0.79	0.19	1	50.068	1.49E-12	0.90	0.624	1	24.63	6.94E-07
TA	0.21	0.81	0	50.068	1.49E-12	0.067	0.281	0	17.23	3.23E-05
TC	0.0					0.016	0.06	0	2.531	0.1117
AA	0.0					0.016	0.06	0	2.531	0.1117

3.3.5 Non-synonymous SNPs and possible mechanisms for their effects on p53 function in *M. edulis*

Three SNPs at positions 219, 261 and 346 were non-synonymous substitutions that were only detected in the diseased phenotypic groups (Table 3.3 Figure 3.5). SNP219 causes substitution of polar acidic serine to non-polar proline; SNP261 changes threonine to serine, both of which are polar, but differ by structure; SNP346 changes proline to leucine, amino acids of the same polarity and very different structures. These variations in amino acid sequence might significantly affect p53 protein secondary structure and, therefore, the activity of the p53 protein.

Table 3.3. Case-control frequencies and significance of association with mussel leukemia for non-synonymous SNPs.

Name	*Association allele*	*Codons*	*Amino acids*	*Haplotype freq. in non-affected group*	*Haplotype freq. in late leukemic group*	*Haplotype freq. in affected group (including all leukemia stages)*
SNP219	C	tcc-ccc	ser-pro	0.000	0.722	0.219
SNP261	T	aca-tca	thr-ser	0.000	0.833	0.328
SNP346	T	ccc-ctc	pro-leu	0.000	0.778	0.344

All three non-synonymous SNPs, 219, 261 and 346 are located in the highly conserved protein region called the proline-rich domain. This domain is responsible for the induction of

cell growth arrest (Ruaro et al. 1997) and initiation of apoptosis (Venot et al. 1998). Both processes are critical for the anti-tumor activity of p53, failure of which leads to the survival and accumulation of neoplastic cells and tumor development. Therefore, at least one or all of these SNPs might have a causative effect on HN in *M. edulis*.

Nevertheless, all detected non-synonymous SNPs have relatively low frequencies, which suggest that these variations are not the only mechanism causing mollusk leukemia. However, since 77.2%, 83.3% and 77.8% of late leukemic animals had these substitutions at positions 219, 261 and 346 respectively, it was proposed that development of the late leukemia stage is associated with the 219C or/and 261T or/and 346T substitutions.

3.3.6 Phylogenetic analysis of *Mytilus trossulus* and *Mytilus edulis* haplotypes

To compare the p53 variable positions between *M. trossulus* and *M. edulis* a phylogenetic analysis was performed on the haplotypes that were predicted using Haploview (Barrett et al. 2005) for the genotypes detected in each *Mytilus* species. Haplotypes were predicted for the 40 variable positions that were found in *M. edulis* p53 gene sequence (Figure 3.2, 3.5). For this analysis 97 *M. trossulus* were analyzed, the same as used for the SNPs detection in Chapter 2, and 58 *M. edulis*, used for the SNPs detection in this Chapter 3. The phylogenetic analysis was done using the Neighbor-Joining method with the Jukes-Cantor model of nucleotide substitution and 1,000 bootstraps using the software package Geneious version 5.4.3 (Figure 3.5). The tree was rooted with the *Danio rerio* mRNA p53 sequence (GenBank accession number NM_001080036 XM_683867).

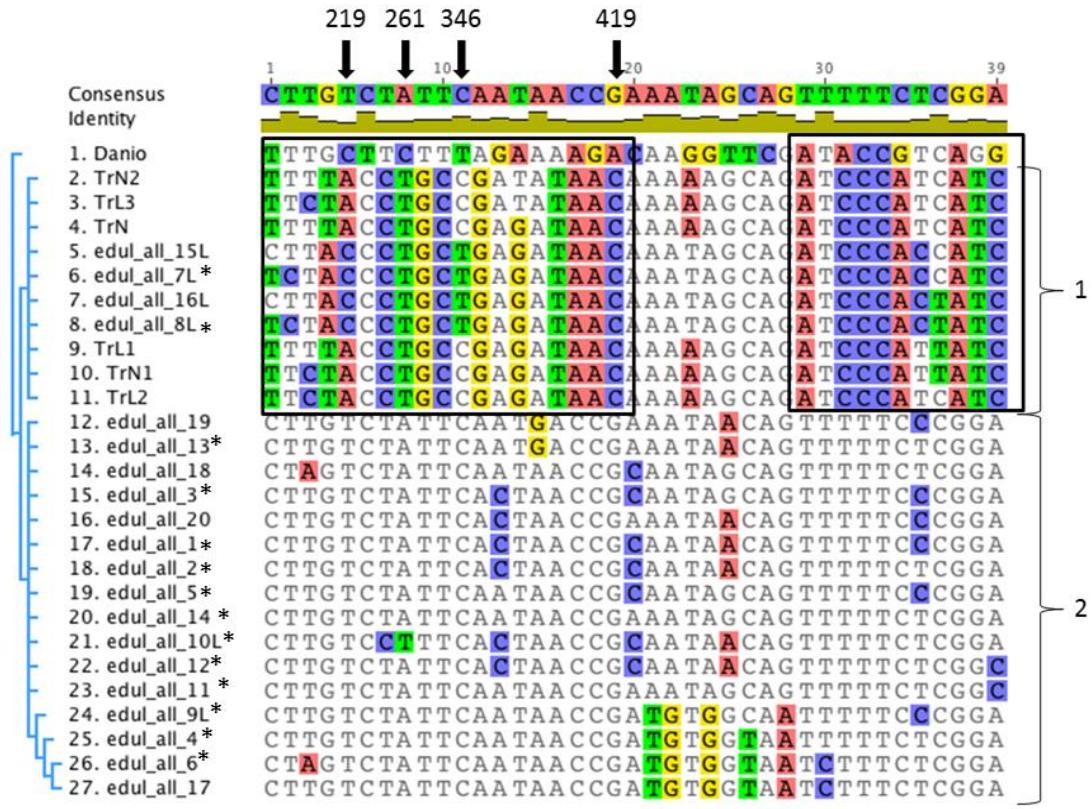


Figure 3.5. Molecular phylogenetic tree for theoretically predicted haplotypes for two mussel species, *M. trossulus* (Tr), and *M. edulis* (Edul). Neighbor-Joining method with the Jukes-Cantor model of nucleotide substitution and 1,000 bootstraps. L: haplotypes predicted for the diseased samples (>10% of neoplastic cells) and which were not detected in normal samples, N –normal *M. trossulus* haplotypes. *: asterisk marks most frequent *M. edulis* haplotypes (that were detected more than once).

All haplotypes predicted for *M. trossulus* are clustered together in one clade regardless of their disease status, while haplotypes predicted for *M. edulis* fell into two clades. Interestingly, the most frequent haplotypes predicted for leukemic *M. edulis* (Edul7 and Edul8, Appendix D) and two rare leukemic haplotypes clustered with the *M. trossulus* haplotypes (Figure 3.4, clade 1), while the most frequent non-diseased *M. edulis* haplotypes (Edul1, Edul2, Appendix D) are distantly related to the *M. trossulus* haplotypes (Figure 3.5, clade 2).

On the haplotype alignment one can see that two blocks of SNPs contribute to the similarity of haplotypes in the clade 1. They are SNPs 1-19 and 29-39 (Figure 3.5, framed). As shown in Figure 3.4c, the two blocks of SNPs in high linkage included SNPs 5 to 19 and SNPs 29 to 37. It is easy to notice that these two pairs of blocks, those clustered with *M. trossulus*

sequences and those in high-linkage, overlap significantly. One possible explanation for this overlapping is hybridization between *M. trossulus* to *M. edulis*, followed by chromosomal recombination. This phenomenon is not unique and was observed for other species (Ruths and Nakhleh 2005).

Further, evidence for the linkage between *M. trossulus* heritage and disease in *M. edulis* comes from comparing Figure 3.4 and Figure 3.5. As seen from the latter, clade 1 contains all *M. trossulus* haplotypes as well as diseased *M. edulis* haplotypes (edul7L, edul8L, edul15L and edul16L). Two sub-blocks of SNPs, 4-19 and 31-39, in clade 1 contain alleles that are in significant association with leukemia: 204A, 219C, 260C, 261T, 267G, 296C, 246T, 380G, 386A, 392G, 401A, 407T, 416A, 419A, 428C (in 4-19 block) followed by 680C, 731C, 737C, 740A, 767C, 821T, 854A 908T, 914C (in 31-39 block). Based on the phylogenetic similarity and on the higher susceptibility to leukemia in *M. trossulus*, it was suggested that *M. edulis* with haplotypes edul7L, edul8L, edul15L and edul16L were descendants of hybrids with *M. trossulus* and that inheriting *M. trossulus*-like haplotypes made them more susceptible to leukemia.

To better illustrate the relationships between these theoretical haplotypes, predicted for *M. trossulus* and *M. edulis*, a haplotype network was generated (Figure 3.6), which was produced using the median-joining network algorithm using the software Network version 4.600 (Bandelt et al. 1999). Two clear clades are formed by the most frequent *M. edulis* and *M. trossulus* haplotypes. An additional group is formed by the four haplotypes, Edul7, Edul8, Edul15 and Edul16, the same ones that clustered with *M. trossulus* haplotypes in Figure 3.5. These four p53 sequence haplotypes are much closer to *M. trossulus* than to *M. edulis* even though all other species markers indicated that these individuals were *M. edulis*.

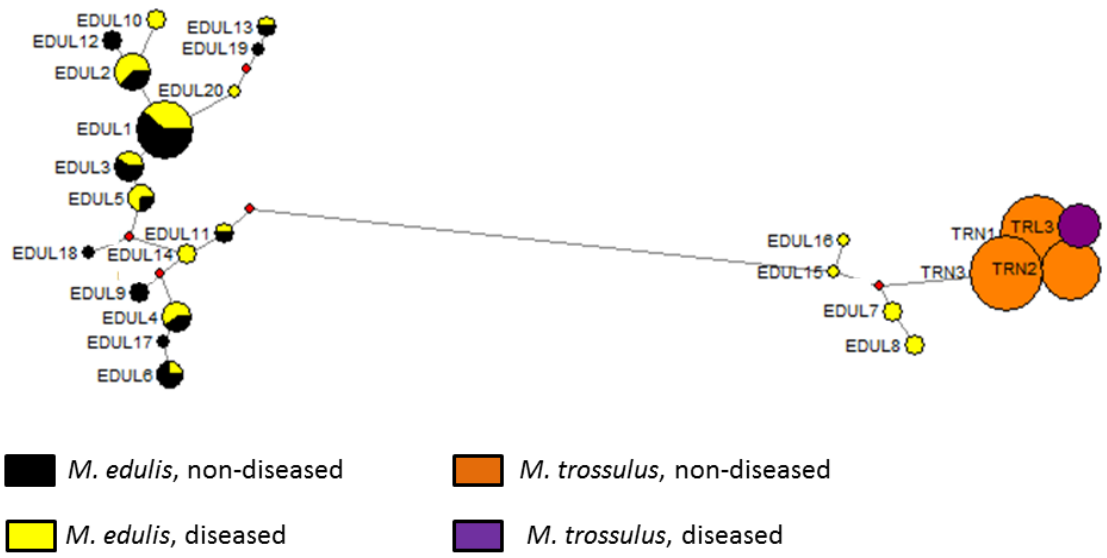


Figure 3.6. Haplotype network for theoretically predicted haplotypes of *M. trossulus* and *M. edulis*. Area of the pies is proportional to the haplotype frequency. Node length is proportional to the amount of substitutions between the neighbors. Numbering of haplotypes is the same as on Figure 3.4. Generated by Network4600 using Median Joining algorithm.

3.3.7 Origin of the detected variations

Phylogenetic analysis of the inferred haplotypes suggested the possibility of hybridization of *Mytilus edulis* with *M. trossulus*, which may have simultaneously introduced susceptibility to leukemia. The three arguments for this hybridization hypothesis and against the polymorphism hypothesis are reiterated: The first one is the remarkably high variability within what is believed to be a highly conserved region. Another argument is that 28 SNPs out of 40 match the species-specific sites, i.e. sites were *M. edulis* p53 sequence differs from the *M. trossulus* p53 sequence. The third one is that all these 28 species-specific SNPs had high frequencies (>0.2) unlike the rest of 12 SNPs with low frequencies (<0.1). Therefore, it was concluded, that such a high number of SNPs within a short fragment with high frequencies of occurrence indicates the possibility of the hybridization. Species identification markers that were used in this study could not detect this hybridization due to the only partial introgression of *M. trossulus* genotype into the *M. edulis* genotype, which affected p53 gene area, but not the areas of genome targeted by the species identification primers.

Mosaic-like structure of such sequence blocks that are in high linkage with each other indicates that the recombination events occurred after hybridization. In other words, the

population that was analyzed contained descendants of the hybrid individual(s). At the same time, no F1-like (first generation of hybrid offspring) genotypes with high numbers of heterozygotes were detected indicating that no recent hybridization events occurred in the population. Hybridization between the two species, *M. edulis* and *M. trossulus*, is limited at prezygotic stage due to the low survivability of the *M. trossulus* trochophores (Toro et al. 2004), which explains the luck of the F1 –like individuals. Hybrid individuals are viable, although they have lower fitness than the pure species. Thus, hybrids and their descendants are capable of maintaining their lineage, some of which were likely present in our population.

3.3.8 Conclusions.

1. Forty variable positions within a fragment of the coding region of p53 mRNA were detected, that formed two blocks of high linkage with each other and that were in strong association with late-stage mussel leukemia.
2. None of the detected SNPs were in significant association with the early stages of the leukemia.
3. Together, a large number of SNPs within a short fragment of a highly conserved gene, high frequencies of these SNPs and LD blocks indicated the likelihood of hybridization between two *Mytilus* species, *M. edulis* and *M. trossulus*, used in this study.
4. The hybridization was not detected by the standard species markers, except for Glu5, probably due to only partial introgression of *M. trossulus* genes into the analyzed *M. edulis* lineage.
5. The non-synonymous SNPs had low frequencies in transitional samples, which suggests the presence of other mechanisms of tumorigeneses.
6. The haplotype LD block pattern and its association with late-stages of the disease might be unique for the population analyzed in this study and cannot be transferred to other populations without a specific analysis.
7. Because the late-stage leukemia *M. edulis* examined in this study had p53 sequences more similar to those for *M. trossulus* it was suggested that *M. edulis*-*M. trossulus* hybrids are more prone to mussel leukemia than pure *M. edulis*. When using mussels for environmental effects monitoring, choosing which species to use is very important due to their differences in disease susceptibility and fitness. In addition, it was shown that even though the commonly used *Mytilus* species-specific genetic markers indicate a particular species, there may be other genetic variabilities arising from hybridization that affect

fitness, and this can confound monitoring studies. When choosing mussels for environmental monitoring more genetic tests should be performed to select a population with an equitable distribution of disease susceptibility among individuals. Also, more genetic tests should be done when selecting which *M. edulis* populations to farm in order to avoid hybrids that may cause increased frequencies of leukemia and high mortality.

4 Flow cytometry as a potential high-throughput technique for detecting leukemia in caged *Mytilus trossulus* populations from Burrard Inlet and off the coast of Hopkins Beach, British Columbia.

4.1 Scope

High throughput and low cost biomarkers are needed for more efficient environmental monitoring and very few are available for sub-lethal endpoints such as leukemia or haemic neoplasia (HN) in mussels. HN results in cell morphology changes and abnormal ploidy in haemocytes. Cell morphology is time consuming and difficult to characterize precisely with a microscope. Flow cytometry (FC) of DNA content was investigated as an alternative, more efficient and possibly more accurate method for detecting HN in mussel populations of two *Mytilus* species, *M. trossulus* and *M. edulis,* from different environments. Results were compared with the conventional hematocytology technique, which estimates the percentage of neoplastic-looking cells in mussel hemolymph. Mussels for this study were collected from beaches and were kept on submerged frames in the Greater Vancouver area (BC, Canada) in 2010. In April, *M. trossulus* were collected from two different beaches and *M. edulis* were purchased from a Vancouver Island shellfish farm. Animals of both species were attached to frames and were submerged in Burrard Inlet from April to October at two different sites, one urban and the other remote. In October, mussels were collected from all frames and additional sampling of beach populations was done. Leukemia prevalence was assessed by hematocytology and hemolymph samples were taken for ploidy analysis.

A new method was developed for analyzing FC ploidy data and for classification of the ploidy patterns. Specific ploidy patterns strongly associated with HN were found for *M. trossulus*, but not for *M. edulis*. The method was used to analyze the beach and caged mussel populations in order to see if there were any differences between them with respect to HN prevalence. Application of the new ploidy method allowed us to show that the natural beach populations were different from caged populations indicating mode of exposure (intertidal or submerged) may confound ecosystem health monitoring. As well, caged populations from the urban exposure site were different from the ones exposed at the remote site, suggesting that the environment conditions may have had an effect on the ploidy status. Overall, ploidy analysis was found to be more sensitive and efficient than hematocytology or genotyping for

monitoring of HN in mussel populations, therefore it was proposed as a potential tool for the ecosystem health monitoring.

4.2 Methods

4.2.1 Mussel exposure experiment

In April 2010, *M. trossulus* longer than 4 cm were collected at low tide from two locations, Hopkins Beach (Sunshine Coast, British Columbia, Canada) and Dufferin Beach (Horseshoe Bay, British Columbia, Canada). *M. edulis* 4.5 to 7 cm long were purchased from the farm Island Scallops Ltd. (Vancouver Island, British Columbia). Forty eight mussels from each beach and sixty mussels from the farm were analyzed to assess the leukemia status using hematocytology as previously described (Vassilenko et al. 2010). Two hundred and forty individuals of *M. trossulus* from each beach and 240 *M. edulis* from the farm were put into net socks, attached to plastic frames (Figure 4.1) and kept submerged at urban (monitoring, M) and remote (reference, R) sites (Figure 2.1) from April until October. Frames were attached to the chain between the sea bottom and the floating boy. The distance from the bottom was fixed and chosen so that the depth from the surface for frames varied from 8 to 12 m depending on the tide level. One hundred and twenty animals from each origin (Hopkins Beach, Dufferin Beach and the shellfish farm, i.e. 360 in total), were used per exposure site (M and R, i.e. 720 grand total). In October, all mussels were removed from the frames and taken to the laboratory for the analysis. At the same week, additional samples were collected from the two beaches and another batch of mussels was received from the farm. The monitoring, or M, site was located in an area that is in the path of effluent from the Lions Gate Wastewater Treatment plant and that is also influenced by water coming out of the Vancouver Harbor. Water samples were collected from both caged exposure sites from the same depth as the frames using a Niskin bottle. Total ammonium and phosphate were measured immediately using CHEMets kits (CHEMetrics, Inc). Metal content was determined using ICP-MS by ALS laboratories (ALS Group, Vancouver, BC). As well fecal coliform analysis was performed by the Metro Vancouver affiliated laboratory.

Figure 4.1. Mussels in net socks, attached to a plastic frame, in preparation for subtidal exposure.

4.2.2 Hemolymph collection and cell staining for flow cytometry

Hemolymph was collected and hematocytological analysis was performed as previously described (Vassilenko et al. 2010), except that this time presence or absence of cell aggregates was recorded in addition to the morphology of individual cells. Briefly, hemolymph was withdrawn from the posterior adductor muscle into a 3 ml syringe preloaded with 0.5 ml of filtered, ice-cold seawater. Haemocyte composition was assessed by examining a drop of hemolymph and samples were classified, as before, into three groups: normal, transitional, and late leukemic regardless of the cell aggregates. Two hundred microliters of hemolymph was fixed in 1300µL of 100% ethanol at room temperature for 30 min and stored at -20^{o}C until further analysis. For flow cytometry, the method described by Delaporte was adopted (Delaporte et al. 2008). On the day of analysis ethanol was removed after centrifugation at 448 × g for 10 min and cells were allowed to rehydrate in 1ml of 0.01M PBS with 2% NaCl at room temperature for 30 min. After another centrifugation at 448 × g for 10 min supernatant was removed and cells were re-suspended in 500µL of 0.01M PBS with 2% NaCl. Cells were transferred into flow cytometry tubes by filtering through an 80 µm nylon mesh to remove all larger particles. 5 µL of DNAse-Free RNase A (Trevigen, 4817-60-04, 20mg/mL) and 10µL propidium iodide (1 mg/mL) were added before incubation for 30 min in dark at room temperature.

Samples were analyzed on a FACSCalibur instrument at the forward scattering (FSC) and fluorescent (FL2) channels. For each run, 10,000 events were counted. The ungated FL2-

H histogram was used to estimate the median DNA content of the sperm population, which was used as a haploid standard. All samples were run and analyzed in sets of 30.

4.2.3 Flow cytometry data analysis.

To analyze the flow cytometry results two gating approaches were applied to each sample. For the first gating method (Figure 4.2 (1a-3a)), the FSC versus FL2-H plot was used to draw regions C1and C2. Region C1 was created based on the normal sample with no additional cells as shown on Panel 1a. This approach excluded the cell debris (particles with low fluorescent signal and small size) in order to analyze only the normal cells (C1) and some other additional (of unknown type) smaller cells (C2) that were detected in 29% of samples (Figure 4.2(2a)). A third gate was set for sperm cells only (Figure 4.2, population "sperm") to assess the fluorescence signal corresponding to haploid cells. At the next step events within the C1 population were plotted as FSC versus FL2-H (Figure 4.2 (3a)) to calculate the percentage of diploid cells (2N), tetraploid cells (4N) and cells with ploidy higher than 4 (nN) in relation to the parental population C1. All gates were set based on a typical normal sample and used for the whole set of 30 regardless of their disease status for the further analysis.

For the second gating method (Figure 4.2 (2b-3b)), the histogram of the fluorescent channel (FL2-H) was used to generate population C (all cells), which gates out cell debris with low fluorescent signal and sperm (Figure 4.2 (1b,2b)). The FL2-H histogram of population C in a typical normal sample was used to set gates for the diploid (2N'), tetraploid (4N') populations and for the population with cells of higher than 4N ploidy (nN') (Figure 4.2 (3b)). Again, these gates were used for all samples within the set of 30.

In order to describe the cell populations in each gate two parameters were used- the percentage of the cells with respect to the parent population and the median of the fluorescence signal (FL2-H) for the population. Population C1 is the parent population for 2N, 4N, and nN populations on the dot plots; population C is the parent population for the 2N', 4N', and nN' on the histograms. For every set of 30 samples the average of the medians for sperm and for normal diploid cell populations were calculated and used it for further analysis.

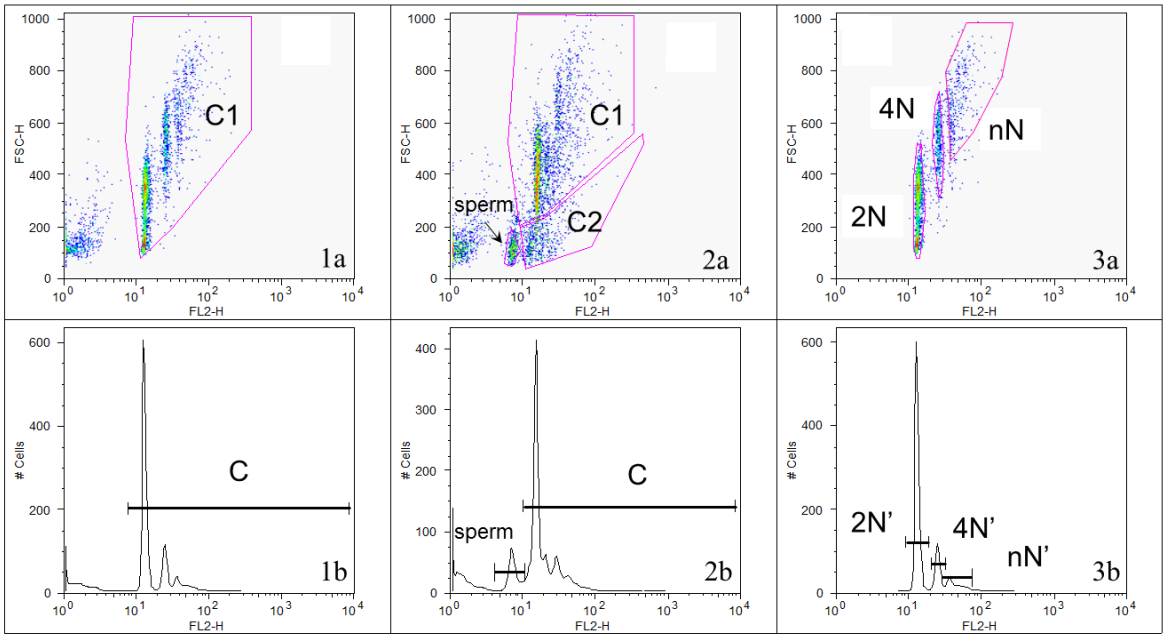


Figure 4.2. Flow cytometry gating methods. Dot plots (1a-3a) show cell size (FSC-H) versus fluorescence (FL2-H) with the color coding for the population density, going along the spectrum from blue for low density to red for high density. Histograms (1b-3b) show cell counts versus fluorescence signal (FL2-H). C1 – gate for normal cells; C2 – gate for additional smaller cells; C – histogram gating for all cells; 2N – gate for the normal diploid population; 4N – gate for the normal tetraploid population; nN – gate for the population of cells with higher than 4 ploidy. 2N', 4N' and nN' gates for the normal populations set on the histogram.

The ploidy of a population was calculated using the fluorescent signal of the sperm as an internal standard as follows:

$$n_i = \frac{N_i}{Nsp}$$

where $n_i$is the ploidy of the population of interest, N_i is the median on the FL2-H axis of population of interest, Nsp is the average of the medians on the FL2-H axis of the sperm populations for the sample set of 30.

Parameter Δ was introduced to calculate the shift between the normal 2N cell population signal and cell populations outside of the pre-set gate for the normal 2N population. In order to describe these populations outside of the normal 2N gates and to calculate their median fluorescence signal an additional gate was set specific for each sample that contained these cells. Parameter Δ was calculated as follows:

$$\Delta = n_i - n_{2N}$$

Where Δ is the ploidy shift, □□ is the ploidy of the outside, or abnormal, population with lowest fluorescent signal (N_i), n_{2N} is the average of the medians on the FL 2-H axis of the normal diploid population for the sample set calculated as :

$$n_{2N} = \frac{\sum_{1}^{30} n_{2N}}{30}$$

4.3 Results

4.3.1 Hematocytology

During the whole project, hematocytology was used to analyze several hundred hemolymph samples. The overall observations made regarding the individual cells are summarized in this section.

According to the literature and as was discussed in Chapter 1 (Section 1.2.3), there are at least three distinct groups of mussel haemocytes: granulocytes, agranulocytes and neoplastic cells (Calisi et al. 2008, Dyrynda et al. 1997, Hine 1999, Pipe et al. 1997, Ratcliffe and Rowley 1981). However according to our observations, hemolymph of both *Mytilus trossulus* and *M. edulis* contained cells with a wide range of morphological features somewhere in between cells with large spread pseudo-podia (1) and smooth round (7), broadly classified in Figure 4.3. For the leukemia status assessment used in this work, cells that looked like 5, 6 and 7 were classified as neoplastic, since these cells lost their adhesion and spreading ability, whereas all others were assumed to be normal cells, adherent and spreadable.

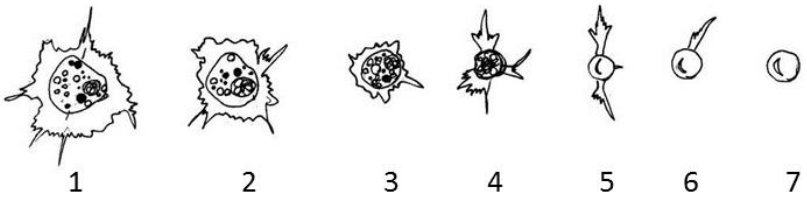


Figure 4.3. Morphological variations of *Mytilus trossulus* and *M. edulis* haemocytes. Left to right: transition between normal and neoplastic haemocytes.

In many samples (up to 70%) haemocyte aggregates were detected in which it was difficult or not possible to recognize the cell types. Features of these cell aggregates and their formation are described in Appendix A. To the best of our knowledge, until now cell aggregates have not been discussed in the context of leukemia assessment. Nevertheless, cell clumping affects the amount of the countable neoplastic cells, present in the hemolymph. Moreover, clumps are present in more than 70% of samples. For example, out of 344 *M. trossulus* analyzed during the exposure experiment in 2010 only 97 samples (or 28%) had no clumps, but only individual haemocytes. It was proposed that clump formation can be prevented by adding anticoagulants (Chen and Bayne 1995) when hematocytology is used as the main method for HN assessment. In this work for leukemia detection only counts of the identifiable cells were used for the analysis.

4.3.2 Normal and abnormal ploidy patterns

To define features of the normal, or "healthy", ploidy pattern, a subset of samples that were assessed as non-diseased by hematocytology was used. In order to decrease the possibility for diseased samples to fall into the non-diseased group by misdiagnosis (false-negative), the threshold level of 10% neoplastic cells per sample was decreased to 5% and samples with large cell aggregates were excluded. All normal samples of both mussel species had similar ploidy patterns such as the one shown on the two-dimensional plots in Figure 4.4 (1a), and on the histograms in Figure 4.4 (1b). Diploid and tetraploid cell populations are clearly distinguishable from other groups, while cells with higher ploidy do not form a distinct population. Boundaries for all three cell populations, 2N, 4N and nN, were set as shown in Figure 4.4(1a). The two-dimensional plots show that cells in the normal 2N and 4N populations are very diverse in size: they vary within 500 FSC-H units out of 1000. The wide range of the haemocyte sizes can be explained by the fact that normal haemocytes vary significantly in function, but all of them maintain the exact number of chromosomes.

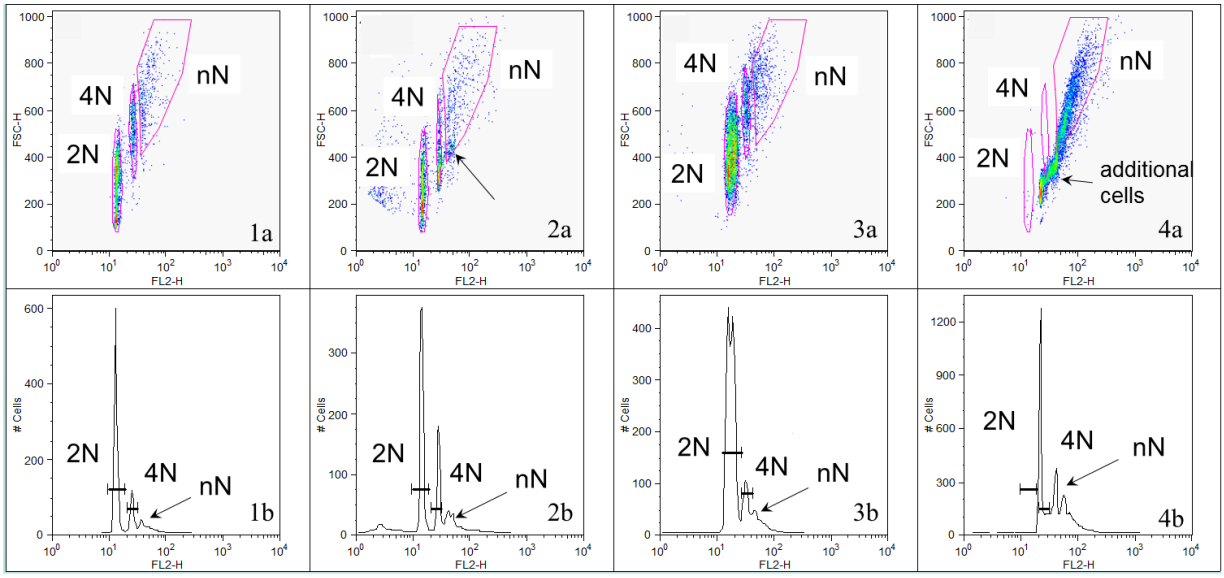


Figure 4.4.Typical ploidy patterns for normal (1a,b) and three diseased (2a,b-4a,b) samples. Panels 2a,b show the transitional sample, where both normal 2N and 4N cells are present, but the amount of 4N cells is increased if compared to the normal sample. Panels 3a,b show transitional sample with the increased amount of 2N cells. Panels 4a,b show late leukemic sample where no normal cells is left.

The percentage of the 2N population over the whole C1 population varied among the normal samples within the range of 40 to 85% for *M. trossulus* and 35 to 95% for *M. edulis*. The relative size of the 4N population varied within 5 to 20% for *M. trossulus* and 0 to 20% for *M. edulis*. The distributions of the relative sizes of 2N and 4N populations among the normal samples (<5% of neoplastic cells) are shown in Figure 4.5. Both the 2N and 4N distributions are more narrow and more even for *M. trossulus* as compared to *M. edulis*.

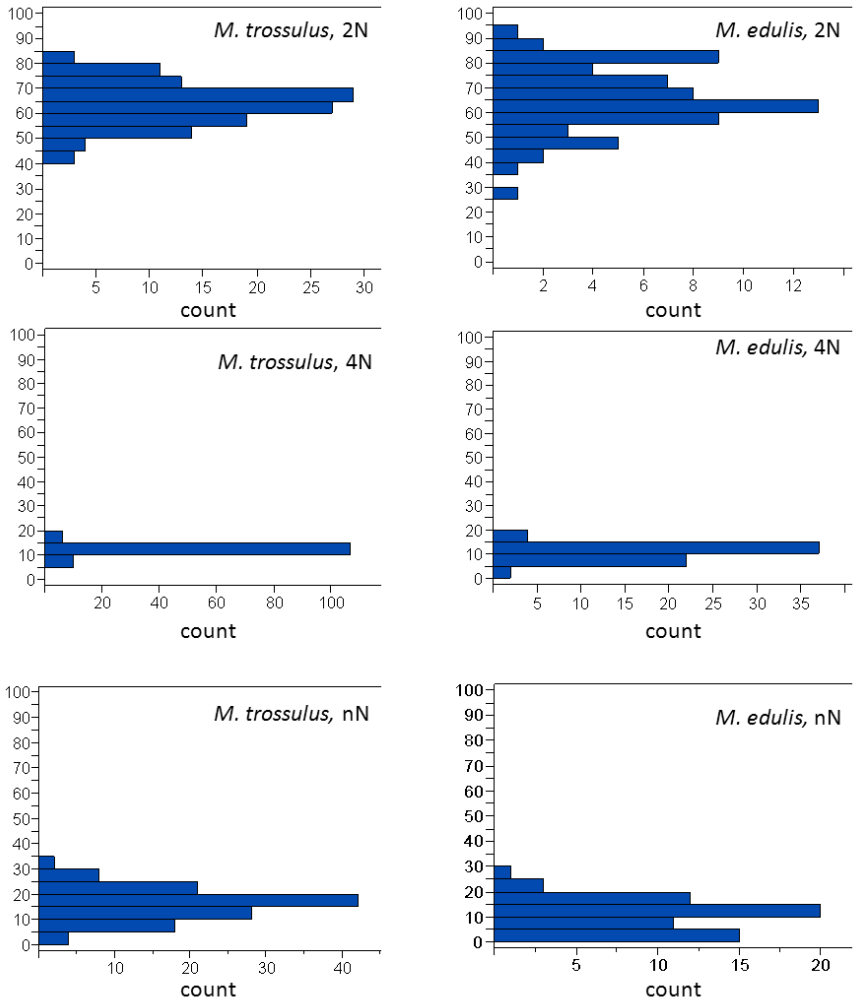


Figure 4.5. Distribution of the relative sizes of diploid (2N), tetraploid (4N) and higher ploidy (nN) populations for non-diseased samples of two *Mytilus* species, *M. trossulus* and *M. edulis*. In each panel, the vertical scale measures the size of the 2N and 4N populations. For each percentage, the bar is proportional to the number of samples in which such a value was found.

In contrast, the ploidy patterns of the transitional and late leukemic hemolymph samples deviated from the "normal" pattern in several ways. These "abnormal" patterns could be divided into two groups. The first group included samples with the overall pattern similar to the normal, but with deviations of the 2N and 4N population sizes beyond the normal range. Range of the normal sizes of 2N, 4N and nN populations is illustrated on the Figure 4.5. Patterns of this type are presented on the Figure 4.4 (2,3), were they have increased 4N and nN populations (panels 2a,b) and increased 2N population (panels 3a,b). The second group of abnormal patterns includes those samples where the ploidy patterns have cell populations other than normal 2N, 4N, and nN. These ploidy patterns showed additional, or abnormal, cell populations, located outside the region of the normal 2N, 4N, and nN populations. Examples of abnormal ploidy patterns of the second kind are given in Figure 4.4

(4a,b). The cell sizes in the abnormal populations vary only within 100 size units for the any particular ploidy value. Whereas, the ploidy values of additional populations vary within a wide range such that the cells were distributed as smears on the two-dimensional plots.

Neoplastic cells lose their ability to differentiate, and therefore they remain one size, unlike normal haemocytes. Neoplastic cell size increases only with the accumulation of DNA as can be seen in the upward trend of the dense smear on the 2-D flow cytometry dot plots. The reason for the appearance of a smear rather than distinct multiples of DNA content could be the randomness of acquisition of extra chromosomes by the neoplastic cells as a result of the cell cycle disruptions, without losing any, as described in (Gonzalez-Tizon et al. 2000).

4.3.3 Haemic Neoplasia types

An earlier study indentified two types of *Mytilus trossulus* leukemia based on DNA content: the pentaploid and tetraploid forms (Moore et al. 1991). Another study found that the karyotype of *Mytilus trossulus* was different for all individual samples analyzed (Gonzalez-Tizon et al. 2000). To clarify this controversy, ploidy of the leukemic haemolymph samples was analyzed. Twenty-three samples of *M. trossulus* with clearly distinguishable abnormal cell populations were found. For these samples the ploidy shifts of the neoplastic cell populations relative to the ploidy of a normal 2N population were calculated. The calculation is described in the "Flow cytometry data analysis" section of Methods. Figure 4.6 shows the ploidy shifts, or the Δ parameter, for all of the 23 samples containing the abnormal cell populations. The ploidy shifts varied between the values -0.2 and 3.0 units for most of samples, and only one sample had an outstanding shift of 5 units.

For *M. edulis*, only 2 samples with leukemic ploidy patterns as described above were found. Therefore, similar analysis for this species could not be performed and any conclusions regarding the ploidy of neoplastic cells could be done.

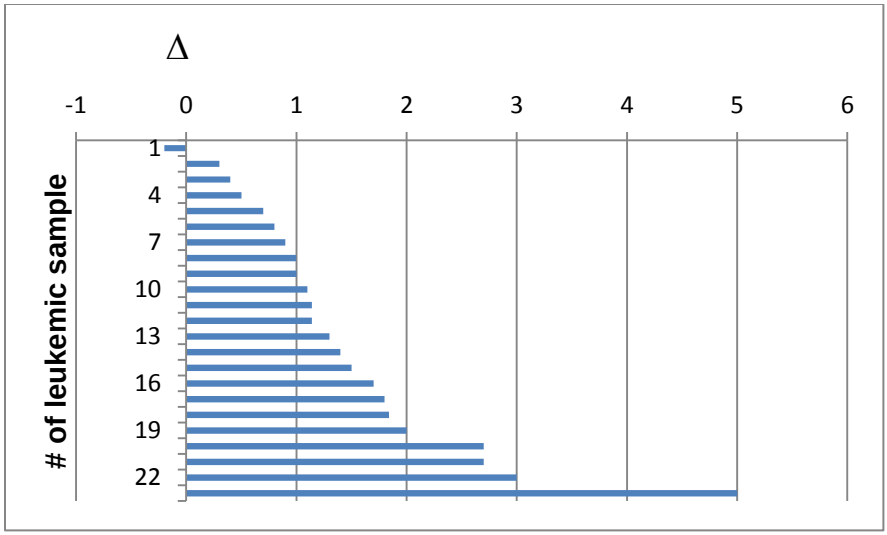


Figure 4.6. Ploidy shifts in abnormal cell populations compared to the normal 2N ploidy population.

4.3.4 Flow cytometry pattern classification: *M. trossulus*

In order to evaluate the possibility of using flow cytometry patterns for detection of all leukemia stages the associations between ploidy patterns and leukemia status was tested as determined by hematocytology. To do so a supervised discriminant analysis was applied, where covariates were the metrics of flow cytometry plots as listed in Table 4.1. The discrimination with the lowest amount of misclassified samples was achieved when the categories were the three leukemic groups described in Table 4.2, which were generated according to the hematocytology. Canonicals 1 and 2 were defined for the linear model for the minimal set of covariates as indicated in Table 4.1. This model grouped all samples as presented in Figure 4.7. All three groups formed distinct clusters, and groups 0 and 1 (representing normal and transitional samples, respectively) were closer to each other than the late leukemic group. This model misclassified 21.7% or 67 out of 309 samples. In contrast to the results above, the ploidy patterns for *M. edulis* did not correlate with leukemia stage-related groups. None of the combinations of ploidy pattern metrics resulted in discrimination of the ploidy patterns into clusters representing leukemic groups. This result is not surprising since the distribution of the sizes of cell populations 2N, 4N and nN (Figure 4.5) were further from normal distribution for *M. edulis* compared to *M. trossulus*. In addition, out of 151 *M. edulis* samples only 2 were late leukemic and had aneuploid patterns distinguishable from the normal ones and similar to presented on Figure 4.4, panel 4a,b. The transitional leukemic samples did not have ploidy patterns statistically different from those for normal animals, according to the pattern-screening algorithm.

Table 4.1. Flow cytometry pattern metrics (as on Figure 4.3) and canonical scores for the discriminant analysis. C - gating of FL2-H histogram, C1-gating of normal cells using two dimensional plot FSC versus FL2-H ; C2 - gating of smaller cells using two dimensional plot FSC versus FL2-H. F ratio shown for the model with four entered covariates.

Gating	*Metrics*	*F Ratio*	*Minimal set of covariates*	*Canonical scores*	
				Canonical 1	*Canonical 2*
	2N'	17.806	x_1	-0.022	0.066
	4N'	0.407			
	nN'	23.650	x_2	0.064	0.033
C1	2N	**311.340**	x_3	0.132	0.033
	4N	0.150			
	nN	35.205	x_4	0.107	-0.053
C2	C2	0.718			

Table 4.2. Leukemic groups generated according to the hematocytology results.

Leukemic groups	*Percent of neoplastic-looking cells, %*
0 Non-diseased	0-10
1 Transitional	10-90
2 Late leukemic	90-100

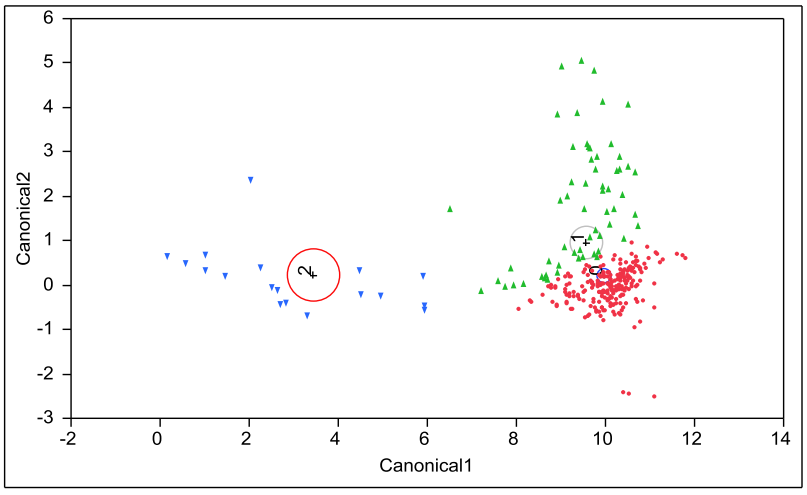


Figure 4.7. Assignment of the *M. trossulus* ploidy patterns into three leukemia groups: 0 (red) –non-diseased; 1 (green) transitional; 2 (blue) late leukemic. The groups were discriminated by the linear model with four covariates (Table 4.1) and tree groups (Table 4.2), using JMP8.

4.3.5 Analysis of mussel populations using flow cytometry and hematocytology

The ultimate goal of the project was to find a method for detecting differences between mussel populations exposed to different environmental conditions. The assignment of the *M. trossulus* samples by flow cytometry to the normal, transitional and leukemic groups as described in the previous section was used to compare eight mussel populations from the 2010 exposure experiment in terms of frequency of the flow cytometry groups.

Mussel populations from two exposure sites (M, urban and R, remote) and two beach sites, Hopkins beach and Horseshoe Bay, (see section 4.2.1) were compared in terms of frequencies of each of the three leukemia groups by hematocytology and flow cytometry (Table 4.3). Cluster analysis was applied to the eight *M. trossulus* populations. Frequencies of leukemic levels were used as covariates. A major difference was detected between one of the April populations (00) and all October populations (**Error! Reference source not found.**) and within the pair of April populations. The distance within the pairs of populations collected from the beaches in October, exposed at M site and at R site were smaller if compared to the distances between the pairs. Based on this it was proposed that season and the exposure mode are the main factor affecting the leukemia prevalence.

For the next step, both April populations were excluded and the discriminant analysis was performed for the October populations only. Frequencies of the leukemia groups were used as covariates, and the exposure mode was used as a grouping parameter. Figure 4.9 illustrates the results of this analysis for the two methods of leukemia assessment, flow cytometry (a) and hematocytology (b).

Discrimination of the six October populations based on the flow cytometry results provides three distinct, not overlapping groups for populations collected at R site (Figure 4.9a group1), M site (group 2) and beach populations (group 0,). Application of discriminant analysis to the hematocytology data does not create three distinct groups, and only R populations can be distinguished from other October populations.

Table 4.3. Frequencies of leukemic groups in *M. trossulus* populations according to flow cytometry and hematocytology. Populations were sampled from two intertidal locations, Hopkins and Horseshoe Bay Beaches in April (Apr) and October (Oct). Intertidal mussels from the same two beaches collected in April were caged until October at two sites, urban (M) and remote (R). Leukemic groups were assigned by hematocytology as described in Table 4.2. Leukemic groups were assigned by flow cytometry based on the discriminant analysis (Figure 4.7).

Population number	*Beach of origin*	*Exposure location*	*Month sampled*	*Frequencies of leukemic groups by flow cytometry*			*Frequencies of leukemic groups by hematocytology*		
				Non-diseased	Transitional	Late leukemic	Non-diseased	Transitional	Late leukemic
1	Hopkins	intertidal	Apr	0.79	0.21	0	0.85	0.15	0
2		intertidal	Oct	0.96	0.02	0.02	0.94	0.04	0.02
3		caged at R	Oct	0.74	0.26	0	0.79	0.21	0
4		caged at M	Oct	0.82	0.13	0.05	0.86	0.11	0.03
5	Horseshoe Bay	intertidal	Apr	0.42	0.54	0.04	0.71	0.25	0.04
6		intertidal	Oct	0.95	0	0.05	0.83	0.1	0.07
7		caged at R	Oct	0.65	0.24	0.11	0.7	0.2	0.11
8		caged at M	Oct	0.74	0.09	0.18	0.71	0.15	0.15

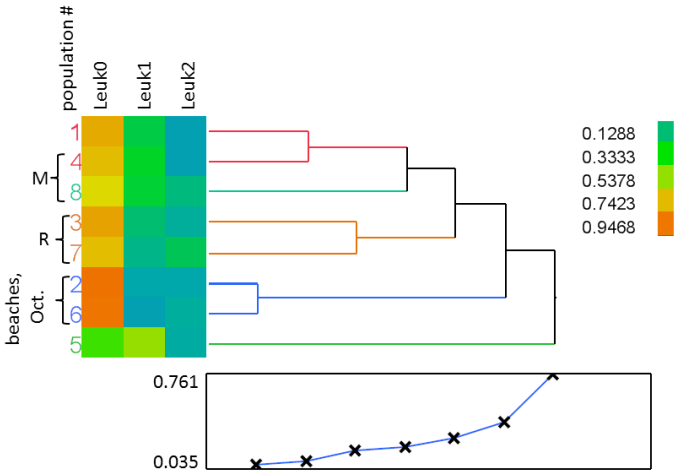


Figure 4.8. Dendogram of eight mussel populations according to the flow cytometry analysis (JMP8). Clustering method is Complete. Color map is generated based on the frequencies of flow cytometry groups and populations are numbered as in Table 4.3. Distance graph (at the bottom) is showing distance overcome at each cluster join.

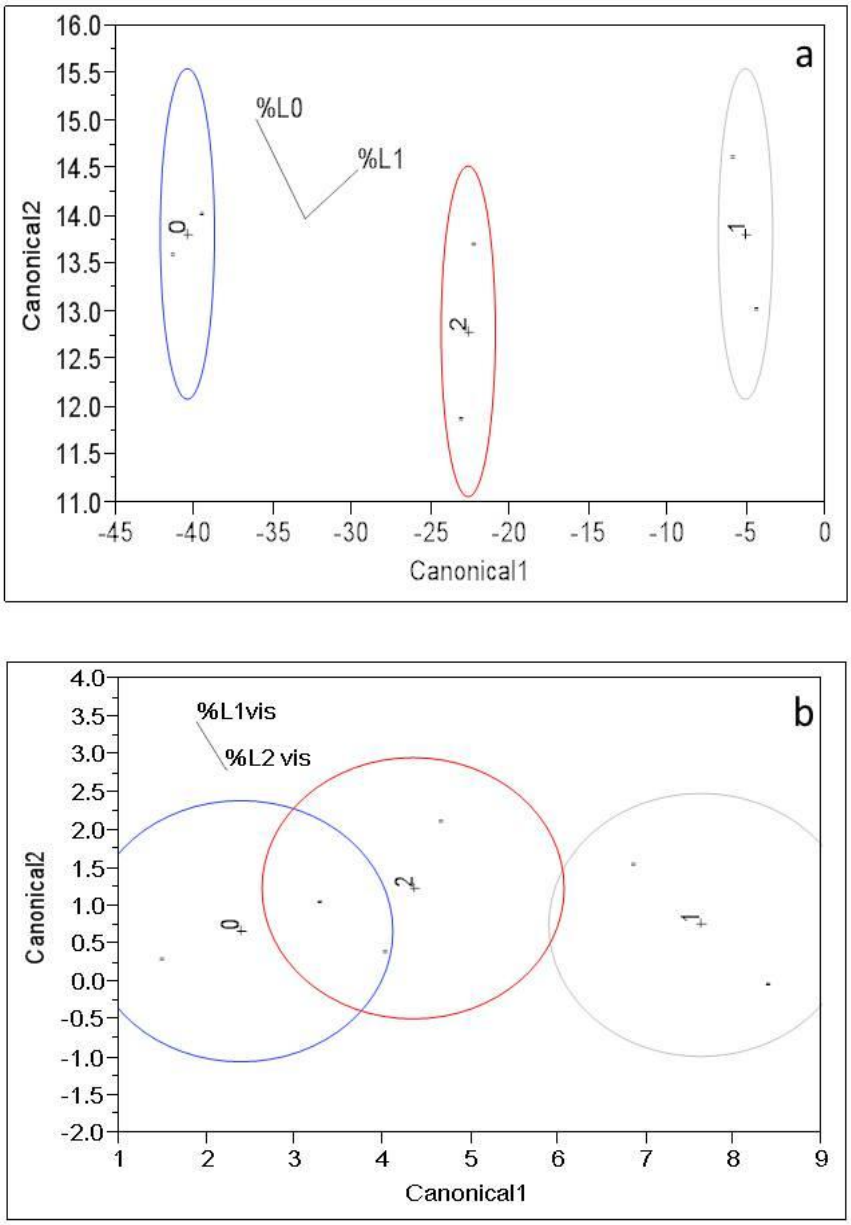


Figure 4.9. Discrimination of populations of *M. trossulus* collected in October based on the exposure mode: 0- non-exposed, collected from the beaches (Hopkins and Horseshoe Bay); 1- exposed on frames at remote sites (R); 2-exposed on frames at urban (M) and collected from two beaches, group 0 based on the leukemia prevalence assessed by flow cytometry (a) and by hematocytology (b). Circles indicate the 95% of confidence limit for the mean.

4.3.6 Water analysis results

Locations of the mussel exposure sites were selected based on the assumption that site M (Figure 2.1) is affected by the effluents of the Lions Gate Wastewater Treatment Plant and Vancouver harbor, while the R site is located in an area of much lower anthropogenic activity. Analysis of water samples collected in April 2010 revealed increased fecal coliforms counts, concentration of ammonia, and concentrations of a number of metals at M site versus R site (Table 4.4). Temperature, pH and dissolved oxygen were similar for both sites. The total ammonia and metals concentrations do not exceed the acceptable level, even according to the most strict standards for drinking water (Health_Canada 2010). Nevertheless, it was shown that some contaminants, such as chlorine and bromine are more toxic in presence of ammonia (Fisher et al. 1999, White 1986). Therefore, even the detected level of ammonia (<0.4mg/L) might have an impact on the mussel health.

Fecal coliform bacteria have been shown to be a good indicator organism for the presence of fecal matter (Bitton 1987, Ericksen 1986), and, therefore, municipal waste water effluent. Fecal coliform counts at the M site were much higher than at the R site (Table 4.4) and than what is acceptable for the safe aquaculture (NSSP 2009). The absolute bacteria counts at M site are two orders of magnitude greater than at the R site, which indicates significantly higher amount of the municipal wastewater effluent, and therefore, higher levels of all contaminants associated with it, such as endocrine-disruption compounds (EDC). Most EDC are known to have a biological effect at very low concentrations, where it is very difficult or impossible to detect them chemically (Falconer 2006). Therefore, it is likely that the environmental effect on the mussel health at the M site is different from the R site, although further work is required to confirm it.

Table 4.4. Results of the contaminant analyses in water samples collected at two exposure sites, remote (R) and urban or monitoring (M). Performed by ALS (8081 Lougheed Highway, Burnaby, Canada). Acceptable concentration limits are presented according to Guidelines for Canadian Water Quality, *: (Health_Canada 2010), **: (CEQG 2010); *** (NSSP 2009).

Sample ID	*R SITE, mg/L*	*M2 SITE, mg/L*	*ACL, mg/L*
Aluminum (Al)-Total	0.022	0.038	0.1-0.2 *
Barium (Ba)-Total	0.0075	0.0096	1*
Boron (B)-Total	3.07	3.15	5*
Calcium (Ca)-Total	249	263	
Lithium (Li)-Total	0.124	0.136	
Magnesium (Mg)-Total	780	831	
Manganese (Mn)-Total	0.00368	0.00528	≤0.05*
Molybdenum (Mo)-Total	0.0068	0.0076	
Potassium (K)-Total	238	248	
Rubidium (Rb)-Total	0.0709	0.0758	
Sodium (Na)-Total	6750	7240	
Strontium (Sr)-Total	4.43	4.77	
Uranium (U)-Total	0.00207	0.00217	0.02*
Vanadium (V)-Total	0.00114	0.00124	
Total ammonia	0-0.2	0.3-0.4	0.7**
Fecal coliforms	0-4	170-460	0*-14***

4.4 Discussion

4.4.1 Ploidy patterns and their association with leukemia

Assessments of leukemia by using hematocytology and flow cytometry were in disagreement for 21.7% of the samples analyzed. In other words, 78.3% of all samples fall into the same leukemic group according to the results of both methods, whereas the rest of the samples were assigned to different groups. This discrepancy may have been caused by two reasons. One reason was that the hematocytological analysis did not accurately diagnose leukemia in some of the samples, which would be especially true for the early stages of the disease where it is difficult to recognize cells with leukemia-related morphological changes.

Another possible reason could have been that leukemia-associated morphological changes in haemocytes, such as the disappearance of pseudopodia or large nucleus to cytoplasm ratios, might not be accompanied by aneuploidy or poly-ploidy since it is unknown which of the events (aneuploidy acquisition or morphological changes) occurs first. More studies are needed to determine at what stage of disease development, aneuploidy or poly-ploidy becomes a measurable feature. The main advantage of flow cytometry over haemocytology is that it is machine-based and allows for more consistent analysis of a larger number of cells from the sample, which is more practical for environmental effects monitoring.

Analysis of the *Mytilus trossulus* late leukemic haemolymph samples did not reveal two distinct groups of ploidy shifts such as the tetraploid and pentaploid forms of leukemia that were described previously by Moore and co-authors (1991). In contrast, a range of aneuploidy was observed comparable to that described by Gonzalez-Trizon et al. (2000) In the *Mytilus trossulus* animals that were studied in this thesis an additional ploidy pattern, not described previously, was observed in the late leukemic samples where a histogram peak was observed near to the diploid DNA content position (2N) on the x-axis but shifted to a ploidy content slightly higher than 2N (Figure 4.4, panel 4b). This finding supports the clonal nature of mussel leukemia. This increase in ploidy content may occur due to cell cycle disruptions in haemocyte precursor that, in turn, gives rise to a whole population of cells of the same karyotype and with abnormal cell cycles. This process is well known for human leukemia and associated with loss of p53 activity (Barlogie et al. 1987, Schmitt et al. 2001). A similar association of aneuploidy with loss of p53 activity may occur in mollusk leukemia.

4.4.2 What factors affect the leukemia prevalence?

DNA ploidy analysis by flow cytometry with the gating methods developed in this thesis was compared with hematocytology by using both procedures to measure incidences of leukemia at two caged study sites and two intertidal areas. The flow cytometry method revealed greater differences in leukemia incidences between the exposure sites than hematocytology. When comparing *M. trossulus* populations from the all the sites using flow cytometry, it appeared that leukemia incidence in the intertidal populations was higher in April than in October which was contrary to the expected increase in the fall (Barber 2004,

Leavitt et al. 1990, Rasmussen 1986) or remaining at a similar level year-round (Brooks and Elston 1991, Farley et al. 1991, McGladdery et al. 2001). The low leukemia prevalence measured in October could have been due to high summer mortalities that would have eliminated the diseased mussels from the population (Bower 1989, Romalde and Barja 2010). This is one of the pitfalls of using a disease such as leukemia as an endpoint, when misleading results may be produced if sampling takes place after large population die-off.

The difference between intertidal and submerged populations was larger than the difference within the group of submerged populations, suggesting that the exposure mode (intertidal or submerged) was one of the main factors affecting leukemia incidence. This might be explained by higher mortality rates among the diseased animals in intertidal areas occurring due to the emersion stresses that are not experienced by caged mussels. Environmental stressors that accompany the emersion stress (UV, extreme temperatures, etc.) may affect disease progression in intertidal mussels populations in other ways, for example exposing mollusk haemocytes to UV in laboratory cultures resulted in overexpression of p53 and initiation of apoptosis in the leukemic cells (Bottger et al. 2008).

The flow cytometry analysis showed that leukemia prevalence differed by location. Of the water quality parameters measured, temperature, pH and dissolved oxygen were similar for both sites, while metal concentrations and fecal coliform counts were higher at the M site. M was an urban site and there may have been many other chemical compounds present that were not measured. For example, surfactants, personal care products and pharmaceutical products are some of the compounds found in wastewater treatment effluent. Also boating activity releases hydrocarbons such as fuel and oils into the ocean. Metal pollution is known to be associated with DNA oxidative damage (Funes et al. 2006), that can affect the leukemia initiation. It has not been shown conclusively that fecal coliforms have a direct effect on mollusk health (Romalde and Barja 2010), but these bacteria are associated with municipal wastewater, that contains various contaminants, including endocrine-disruption compounds (EDC) (Depledge and Billinghurst 1999). EDCs affect overall mollusk health (Depledge and Billinghurst 1999, Porte et al. 2006) and a strong association between EDC exposure and tumorigenesis is well known (reviewed by (Soto and Sonnenschein 2010)). There are many more unaccounted for factors that might be different between the M and R sites, such as food

availability. To learn more about which specific factors could affect leukemia prevalence, controlled exposure experiments would be needed.

Based on this comparison of leukemia prevalence in *M. trossulus* collected from different sites, the following recommendations were proposed. (a) Mussels used for monitoring on submerged frames at the monitoring site must be compared with a reference population, also submerged on frames, at a pristine site during the same time period using mussels originating from the same wild population. This is so that the study site and reference site mussels have the same leukemia prevalence at the beginning of the experiment. Exposing mussels at the sites over the same time period takes into account leukemia prevalence variations that can occur at different times of the year (Barber 2004). (b) Mussels should be sampled simultaneously from the study and reference sites, at least within a week, since changes in environmental conditions could occur affecting leukemia prevalence. (c) The health status of mussels exposed on submerged frames cannot be compared with that of intertidal mussels, even if they are in close proximity, since each population is exposed to different environmental conditions including those that can affect mortality or leukemia prevalence.

5 Conclusions

The thesis aimed to develop new biomarkers that are more accurate and efficient than microscopy-based hematocytology for assessment of mussel haemocyte leukemia, which is a sub-lethal endpoint used in population health monitoring. The specific new biomarkers were p53 gene- and DNA ploidy-based due to previously shown associations between these factors and mussel leukemia. Two mussel species, *Mytilus trossulus* and *Mytilus edulis*, commonly used in environmental affects monitoring, were used. *M. trossulus* were collected from intertidal populations at several beaches in the B.C. Lower Mainland area and *M.edulis* were purchased from the farm. Both species had been caged and submerged for several months from buoys deployed at urban sites near West Vancouver and remote sites near Hopkins Beach, The Sunshine Coast. The working hypotheses were that p53 mutants or sequence variants and increases in DNA ploidy were associated with progression of mussel leukemia and therefore could be used as biomarkers for the disease.

Sequence variations in the p53 mRNA were found that were in association with leukemia for both *M. trossulus* and *M. edulis*. In *M. trossulus,* the association of heterozygote CT at position 821 (genotype C821T) with the disease was valid for mussel populations exposed on the submerged frames but not for the intertidal populations. All mussels collected from the particular intertidal populations sampled in this study were heterozygous (C821T) regardless of their disease status. It was hypothesized that factors affecting p53 gene expression or leukemia are different for intertidal mussels and for caged mussels that are always submerged below the ocean surface, however further investigation is required to identify these factors. Some examples of stressors experienced by intertidal mussels and not those on the cages are UV exposure, desiccation, and high temperatures. These stressors may contribute to higher mortality rates for intertidal leukemic, therefore mussels caged on the submerged frames are more suitable for environmental effects monitoring.

The findings of this thesis contribute to the understanding of processes associated with mussel leukemia. Messenger-RNA sequence variations in the coding region of p53 in *M. trossulus* that were associated with the disease in caged mussels indicated that the development of leukemia in mussels could be regulated not only at the protein level, but also

at the DNA and/or RNA levels. The sequence variations that were detected might be acquired either as a result of a mutation process or due to differential expression of gene alleles. The presence of the "disease-associated" mRNA p53 sequence variants in intertidal mussels with seemingly healthy hemolymph indicated that exposure mode (intertidal or submerged) might also affect gene expression and intertidal mussels may not be suitable for environmental monitoring for this reason..

In *M. edulis,* leukemia-associated-SNPs were detected in individuals with p53 sequences more similar to that for *M. trossulus* than healthy *M. edulis*. This observation lead to the hypothesis that hybridization may have occurred in the founder populations, followed by recombination, creating a mosaic structure of high linkage disequilibrium blocks in the gene of the hybrid. This mosaic structure detected in leukemic farmed *M. edulis* from the frames submerged in Burrard Inlet might be unique for this particular population only. Several other SNPs were detected in the *M. edulis* animals that did not display the hybrid genotype, but none of these had a significant association with leukemia. It is known that *M. trossulus* is more susceptible to leukemia than *M. edulis*. Therefore, it was postulated that the proposed hybridization with *M. trossulus* caused an increase in susceptibility to leukemia for these particular animals. Therefore, if *M. edulis* were to be used for environmental monitoring it will be important to test the founder population for hybrids, including the p53 gene in the suite of hybrid biomarkers, so as to remove inter-individual variability in disease susceptibility. Low susceptibility of *M. edulis* to leukemia makes the species unsuitable as an indicator organism when this endpoint is included into the environmental monitoring program. However this issue might be specific for the Pacific coast, where *M. edulis* is non-endemic species.

Abnormal ploidy was associated with *M. trossulus* leukemia and a new method of ploidy analysis using flow cytometry was developed. Unlike other studies (Moore et al. 1991), where two distinct disease types with tetra- and penta-ploidy were detected in *M. trossulus* only one type of ploidy pattern associated with leukemia was observed, where ploidy varied over a wide range from 1.8 to 7 (where 1 would be normal haploid type). This disagreement with the previous study may be due to the fact that the detected ploidy patterns are population-specific and/or due to the difference in the data analysis methods. In our study intertidal *M. trossulus* from two different beaches, Hopkins Beach and Horseshoe Bay Beach, had the same ploidy

patterns. Association of flow cytometry patterns with leukemia stages was consistent for all the populations of *M. trossulus*, used in this study, i.e. populations caged at M and R sites and intertidal populations collected from Hopkins and Horseshoe Bay beaches in April and October. Therefore it was proposed that ploidy patterns are not population specific for the B.C. Lower Mainland region areas sampled in this study, although more populations need to be analyzed to confirm this suggestion.

For the data analysis mussel sperm was used as an internal standard of haploid signal to calculate the ploidy of other peaks, which compensates for the intersample variations caused by the staining inconsistency. In addition two-dimensional plots (cell size versus fluorescent signal) as wells as histograms were employed to set gates for cell populations in the current study versus only histograms, which allows to differentiate between cells of normal size and smaller ones. It was proposed that both of these measures increase the accuracy of the pattern classification, and therefore, sensitivity of the method.

A set of metrics determined from the flow cytometry pattern for each hemolymph sample was used to assign the sample to one of three groups: normal, transitional, or late leukemic. Based on this classification method, frequencies of the three groups in *M. trossulus* populations exposed at the urban site M were different from those observed in the populations exposed at the remote site R.

Environmental conditions at the two exposure sites, M and R, could be different in many ways. Of the several water quality parameters measured, the presence of higher fecal coliform counts at site M was the most significant difference. Therefore, water quality may be one of the factors affecting mussel haemocyte ploidy, therefore controlled exposure experiments, either on-site or in a laboratory, are required to confirm this hypothesis.

Processing of samples for flow cytometry required less steps, was more robust, and required less operator training than genotyping. Unlike microscopy, the flow cytometry analysis is machine-based, therefore it is unbiased and fast. The resolution of the flow cytometry for population comparison was higher than for microscopy, which allowed distinguishing the populations from different exposure sites.

Altogether, these features make the method of ploidy status assessment an efficient and accurate tool for *M. trossulus* population health monitoring. Nevertheless, the exposure experiment design had a number of limitations. The exposure sites are located far away from each other, M site deep in the Burrard Inlet, and R site is close to the Georgia strait area. Therefore, there might be several uncontrolled parameters affecting the leukemia level in population, such as availability and quality of food or amplitude of variations of temperature, pH, dissolved oxygen concentration etc. In addition, only two populations per exposure site were used, which originated from two different beach populations. Therefore to confirm association of ploidy variations and water quality at the exposure sites more populations per exposure site should be analyzed. As well, additional work is required to verify the applicability of the ploidy marker to mussel populations exposed at additional urban and remote sites.

The importance of the including of mussel leukemia into the biomonitoring programs originates from the growing evidence of the importance of the combined risk assessment (Silins and Hogberg 2011) since mussel populations are exposed to complex stress, the effect of which is difficult to predict and assess. Mussel leukemia should be combined with other end points for mussel population monitoring so all together they cover a wide range of stress responses, including response to chemical carcinogenesis and EDC effect. In addition, single marker information does not provide the opportunity to restore the exposure data (Georgopoulos 2009, Mosquin 2009). Therefore markers should include population level markers, such as growth and reproductive rate, cellular stress response markers such as lysosome stability, molecular markers, such as DNA integrity and protein activity, such as Cytochrome or Heat Shock Proteins activity. The combination of markers of exposure and response at different biological levels will provide comprehensive data on the mussel populations health as well as on specific stressors present in the environment and potentially suggesting the improvement of water management approach.

6 Future work

1. Further studies are needed to verify that polyploidy, which was proposed as a leukemia marker in Chapter 4, is associated with water quality differences at the two exposure sites. Polyploidy and leukemia prevalence in mussel populations might be affected by numerous factors and it is difficult to account for all of them. This study showed that exposure mode, intertidal or submerged, affected the leukemia levels, however other parameters, such as temperature, salinity, food availability and quality, oxygen concentration, pH, viruses, ultra violet light etc. might be important. Therefore, in order to investigate if water quality does affect leukemia prevalence, controlled laboratory exposure experiments are needed where only water quality parameter is varied. For example, in one such experiment *M. trossulus* collected from wild populations could be placed on frames and exposed in flow-through tanks to seawater only (the control) and to sea water to which an effluent, such as the discharge from a Wastewater Treatment Plant, has been added at environmentally relevant concentrations. The initial and final levels of leukemia in mussel populations can be assessed using the techniques developed in this thesis, namely flow cytometry. This study would need to use mussels from several original locations, and even may need to be repeated in different seasons since conditions at the site of origin at the time of mussel collection may confound the results.

2. Municipal effluents contain many different chemical and biological constituents, however, in general are characterized by having high concentrations of pathogens. To explore if these pathogens are involved in disease development, other processes associated with exposure to wastewater plant discharges and mussel leukemia development can be measured. For example, screening for expression of stress-response genes using microarray technologies could reveal information about the molecular pathways that are affected by water quality and potentially allow to distinguish between water quality effect and effects of other confounding factors.

3. For use in aquaculture, *M. edulis* species have a number of advantages over *M. trossulus* or *M. trossulus-edulis* hybrids, such as high growth rate, low leukemia susceptibility and mortality. Therefore, avoiding *edulis-trossulus* hybrids in original mussel stock will increase efficiency of mussel production. Variations in *M. edulis* p53 coding region,

associated with the increased leukemia frequency and indicating the hybridization, will increase sensitivity of species- identification markers developed so far and described in Chapter 4. However, to make the method cost and labor efficient a restriction length polymorphism marker should be developed for *M. edulis* p53 variations. Our preliminary analysis showed that out of the two LD blocks, described in Chapter 3, the first one is located in one exon. Therefore DNA extraction and direct PCR can be used instead of the expensive and tedious mRNA manipulations. A set of restrictases which would produce a species-specific DNA fragment pattern for *Mytilus* need to be determined and verified by sequence analysis of PCR amplified p53 fragment.

4. The decreased amount of heterozygotes in *M. trossulus* populations from the submerged frames, presented in Appendix B, might be caused by the differential expression of p53 family from different alleles in response to the extreme conditions during the emersion. mRNA sequence analysis of p53 isoforms, expressed under different conditions, intertidal and submerged, is required to confirm our hypothesis. For this analysis p53 isoforms should be amplified with isoform-specific primers described in Chapter 2. The three variable positions detected for *M. trossulus* can be analyzed using sequencing as it has been done before or using High Resolution Melting Analysis of PCR products as described in Appendix E. Alternatively, next generation Pyrosequencing of p53 isoforms might be performed, which would significantly increase the high throughput and cost efficiency of the analysis (Barbazuk and Schnable 2011).

References

Ainsworth, T. D., K. Wasmund, L. Ukani, F. Seneca, D. Yellowlees, D. Miller, and W. Leggat. 2011. *Defining the tipping point. A complex cellular life/death balance in corals in response to stress*. Macmillan Publishers Limited.

Almoguera, C., A. Rojas, and J. Jordano. 2002. Reversible Heat-Induced Inactivation of Chimeric Glucuronidase in Transgenic Plants. *Plant Physiol.* 129: 333–341.

Anestis, A., A. Lazou, H. O. Pörtner, and B. Michaelidis. 2007. Behavioral, metabolic, and molecular stress responses of marine bivalve *Mytilus galloprovincialis* during long-term acclimation at increasing ambient temperature. *Am J Physiol Regul Integr Comp Physiol* 293: R911–R921.

Armenian, H. 1987. Incubation periods of cancer: old and new. *J Chronic Dis.* 40: 9S-15S.

Arrowsmith, C. 1999. Structure and function in the p53 family. *Cell Death & Diff.* 6: 1169 ± 1173.

Arrowsmith, C. H. 2003. Structure and function in the p53 family. *Cell Death & Diff.* 6: 1169-1173.

Ashby, J., and R. W. Tennant. 1991. Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP. *Mutat Res-Genet Tox* 257: 229-306.

Babendure, J. R., J. L. Babendure, and J.-H. Ding. 2009. Control of mammalian translation by mRNA structure near caps. *RNA* 12: 851–861.

Baldwin, J., and P. W. Hochachka. 1970. Functional significance of isoenzymes in thermal acclimatization. Acetylcholinesterase from trout brain. *Biochem J.* 116: 883-887.

Balouet, G., M. Poder, A. Cahour, and M. Auffret. 1986. Proliferative hemocytic condition in European flat oysters (*Ostrea edulis*) from Breton coasts: A 6-year survey. *J Invertebr Pathol.* 48: 208-215.

Bandelt, H., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16: 37-48.

Banni, M., Negri, A, Rebelo, M, Rapallo, F, Boussetta, H, Viarengo, A, Dondero, F. 2009. Expression analysis of the molluscan p53 protein family mRNA in mussels (*Mytilus* spp.) exposed to organic contaminants. *Comp Biochem Physiol C Toxicol Pharmacol.* 149: 414-8.

Barbazuk, W., and P. Schnable. 2011. SNP discovery by transcriptome pyrosequencing. *Methods Mol Biol* 729.

Barber, B. J. 2004. Neoplastic diseases of commercially important marine bivalves. *Aquat. Living Resour.* 17: 449-466.

Barker, C. M., R. J. Calvert, C. W. Walker, and C. L. Reinisch. 1997. Detection of mutant p53 in clam leukemia cells. *Exp. Cell Res.* 232: 240-245.

Barlogie, B., S. Stass, D. Dixon, M. Keating, A. Cork, J. Trujillo, K. McCredie, and E. Freireich. 1987. DNA aneuploidy in adult acute leukemia. *Cancer Genet Cytogenet* 28: 213-28.

Barrett, J. C., B. Fry, J. Maller, and M. J. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263-265.

Baumann, P. C. 1998. Epizootics of cancer in fish associated with genotoxins in sediment and water. *Mutat Res-Genet Tox* 411: 227-233.

Baumann, P. C.. 2010. Data Analysis and Fish Tumor BUI Assessment For the Lower Great Lakes and Interconnecting Waterways. Environment Canada.

Becker, K., P. Pancoska, N. Concin, K. V. Heuvel, N. Slade, M. Fischer, E. Chalas, and U. M. Moll. 2006. Patterns of p73 N-terminal isoform expression and p53 status have prognostic value in gynecological cancers. *Int. J. Oncol.* 29: 889-902.

Belyi, V. A., P. Ak, E. Markert, H. Wang, W.Hu, A. Puzio-Kuter, and A. J. Levine. 2010. The origins and evolution of the p53 family of genes. *Cold Spring Harb. Perspect. Biol.*: 1-17.

Benard, J., S. Douc-Rasy, and J.-C. Ahomadegbe. 2003. Tp53 family members and human cancers. *Hum. Mutat.* 21: 182-191.

Benfenati, E., Benigni, R, Demarini, DM, Helma ,C, Kirkland, D, Martin, TM, Mazzatorta, P, Ouédraogo-Arras, G, Richard, AM, Schilter, B, Schoonen, WG, Snyder, RD, Yang, C.,. 2009. Predictive models for carcinogenicity and mutagenicity: frameworks, state-of-the-art, and perspectives. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev.* 27: 57-90.

Beranek, D. T. 1990. Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. *Mutat Res-Fund Mol M* 231: 11-30.

Bernstein, C. 2009. DNA Damage and Cancer. *SciTopics.*

Bihari, N., M. Micic', R. Batel, and R. Zahn. 2003. Flow cytometric detection of DNA cell cycle alterations in hemocytes of mussels (*Mytilus galloprovincialis*) off the Adriatic coast, Croatia. *Aquat Toxicol.* 64: 121-129.

Bitton, G. 1987. Fate of bacteriophages in water and wastewater treatment plants. Pages 181-195 in C. P. G. S.M. Goyal, G. Bitton, ed. *Phage Ecology*. Wiley Interscience, New York.

Borthagaray, A. I., and A. Carranza. 2007. Mussels as ecosystem engineers: Their contribution to species richness in a rocky littoral community. *Acta oecologica* 3 1: 2 4 3 – 2 5 0.

Bottger, S., E. Jerszyk, B. Low, and C. Walker. 2008. Genotoxic stress-induced expression of p53 and apoptosis in leukemic clam hemocytes with cytoplasmically sequestered p53. *Cancer Res* 68: 777-782.

Bourdon, J.-C., K. Fernandes, F. Murray-Zmijewski, G. Liu, A. Diot, D. Xirodimas, M. Saville, and D. Lane. 2005. p53 isoforms can regulate p53 transcriptional activity. *Genes Dev.* 19: 2122-2137.

Bower, S. M. 1989. The summer mortality syndrome and haemic neoplasia in the blue mussels (*Mytilus edulis*) from British Columbia. *Canadian Technical Report of Fisheries and Aquatic Sciences* 1703: 1-65.

Bower, S. M. 2006. Synopsis of infectious diseases and parasites of commercially exploited shellfish: Haemocytic neoplasia of mussels.

BRER. 2006. *Radiation-Induced Cancer: Mechanisms, Quantitative Experimental Studies, and the Role of Genetic Factors.*

Bresler, V., O. Mokady, L. Fishelson, T. Feldstein, and A. Abelson. 2003. Marine molluscs in environmental monitoring II. Experimental exposure to selected pollutants. *Helgol Mar Res* 57: 206-211.

Brooks, K., and R. A. Elston. 1991. Epizootiology of haemic neoplasia in *Mytilus trossulus* in Washington State. II. *J. Shellfish Res.* 10: 233.

Brown, R., Wolke, RE, Saila, SB, Brown, CW. 1978. Prevalence of neoplasia in 10 New England populations of the soft-shell clam (*Mya arenaria*). *Ann N Y Acad Sci.* 298: 522-34.

Brown, R. S. 1980. The value of the multidisciplinary approach to research on marine pollution effects as evidenced in a three-year study to determine the etiology and pathogenesis of neoplasia in the soft-shell clam, *Mya arenaria*. *Rapports et Proces-*

Verbaux des Reunions, Conseil International pour L'Exploration scientifique de la Mer Medeterranee. Monaco 179: 125-128.

Buesching, C., M. Heistermann, and D. Macdonald. 2009. Seasonal and inter-individual variation in testosterone levels in badgers: evidence for the existence of two endocrinological phenotypes. Pages 865-871. *J Comp Physiol A: Neuroethology, Sensory, Neural, and Behavioral Physiology*. Springer Berlin / Heidelberg.

Burga, C. A., and A. Kratochwil. 2001. *Biomonitoring: General and Applied Aspects on Regional and Global Scales*. Kluwer Academic Publishers.

Calisi, A., M. Lionetto, R. Caricato, M. Giordano, and T. Schettino. 2008. Morphometric alterations in *Mytilus galloprovincialis* granulocytes: a new biomarker. *Environ. Toxicol. Chem.* 27: 1435-1441.

Canova, S., Degan, P, Peters, LD, Livingstone, DR, Voltan, R, Venier, P. 1998. Tissue dose, DNA adducts, oxidative DNA damage and CYP1A-immunopositive proteins in mussels exposed to waterborne benzo[a]pyrene. *Mutat Res.* 399: 17-30.

Cantillo, A. 1991. National Status and Trends Program for Marine Environmental Quality. Mussel Watch Worldwide Literature Survey. *NOAA Technical Memorandum*.

Carballal, M., M. Lopez, C. Azevedo, and A. Villalba. 1997. Hemolymph cell types of the mussel *Mytilus galloprovincialis*. *Dis. Aquat. Org.* 29: 127-135.

Carlson, C. S., M. A. Eberle, M. J. Rieder, Q. Yi, L. Kruglyak, and D. A. Nickerson. 2004. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet.* 74: 106-20.

Carroll, S., J. Grenier, and S. Weatherbee. 2004. *From DNA to diversity: molecular genetics and the evolution of animal design.* Blackwell Publishing Ltd, Oxford, UK.

Casanueva, M. O., A. Burga, and B. Lehner. 2012. Fitness Trade-Offs and Environmentally Induced Mutation Buffering in Isogenic C. elegans. *Science* 335: 82-85.

Castano-Vinyals, G., A. D'Errico, N. Malats, and M. Kogevinas. 2004. Biomarkers of exposure to polycyclic aromatic hydrocarbons from environmental air pollution. *Occup Environ Med.* 61.

CEQG. 2010. Canadian Water Quality Guidelines for the Protection of Aquatic Life. Ammonia. *Canadian Environmental Quality Guidelines; Canadian Council of Ministers of the Environment*.

Cesková, P., Valík, D, Vojtesek, B. 2003. What we currently know about the structure and function of the p53 homologue - p73 protein: facts, hypotheses and expectations. *Folia Biol (Praha).* 49: 1-8.

Cetin-Atalay, R., and M. Ozturk. 2000. p53 mutations as fingerprints of environmental carcinogens. *Pure Appl. Chem.* 72: 995-999.

Chan, H. 1988. Accumulation and tolerance to cadmium, copper, lead and zinc by the green mussel *Perna viridis*. *Mar. Ecol. Prog. Ser.* 48: 295-303.

Chao, H., K. Hsiao, and T. Su. 2001. A silent mutation induces exon skipping in the phenylalanine hydroxylase gene in phenylketonuria. *Hum. Genet.* 108: 14-19.

Chen, F., Chang, D, Goh, M, Klibanov, SA, Ljungman, M. 2000. Role of p53 in cell cycle regulation and apoptosis following exposure to proteasome inhibitors. *Cell Growth Differ.* 11: 239-46.

Chen, J.-H., and C. J. Bayne. 1995. Bivalve Mollusc Hemocyte Behaviors: Characterization of Hemocyte Aggregation and Adhesion and Their Inhibition in the California Mussel (*Mytilus californianus*). *Biol Bul* 188: 255-266.

Chi, S., A. Ayed, and C. Arrowsmith. 1999. Solution structure of a conserved C-terminal domain of p73 with structural homology to the SAM domain. *The EMBO Journal* 18: 4438 - 4445.

Ching, E. W., W. Siu, P. Lam, L. Xu, Y. Zhang, B. J. Richardson, and R. Wu. 2001. DNA adduct formation and DNA strand breaks in green-lipped mussels *(Perna viridis*) exposed to benzo[a]pyrene: dose- and time-dependent relationships. *Mar Pollut Bull* 42: 603-10.

Cho, Y.-C., R. C. Frohnhoefer, and G.-Y. Rhee. 2004. Bioconcentration and redeposition of polychlorinated biphenyls by zebra mussels (*Dreissena polymorpha*) in the Hudson River. *Water Res* 38: 769-777.

Ciocan, C. M., and J. M. Rotchell. 2005. Concervation of cancer genes in the marine invertebrate *Mytilus edulis*. *Environ Sci Technol* 39: 3029-3033.

Collins, K., T. Jacks, and N. P. Pavletich. 1997. The cell cycle and cancer. *PNAS* 94: 2776-2778.

Comesana, A. S., J. E. Toro, D. J. Innes, and R. J. Thompson. 1999. A molecular approach to the ecology of a mussel (*Mytilus edulis - M. trossulus*) hybrid zone on the east coast of Newfoundland, Canada. *Mar Biol* 133: 213-221.

Connell, D. W., P. K. S. Lam, B. J. Richardson, and R. S. S. Wu. 1999. *Introduction to ecotoxicology.* Oxford, UK ; Malden, MA, USA : Blackwell Science.

Cortazzo, P., C. Cerveñansky, M. Marín, C. Reiss, R. Ehrlich, and A. Deana. 2002. Silent mutations affect in vivo protein folding in Escherichia coli. *Biochem. Biophys. Res. Commun.* 293: 537-541.

Cosson, R. 2000. Bivalve metallothionein as a biomarker of aquatic ecosystem pollution by trace metals: limits and perspectives. *Cell Mol Biol* 46: 295-3009.

Cox, R. L., R. E. Stephens, and C. L. Reinisch. 2003. p63/73 homologues in surf clam: novel signaling motifs and implications for control of expression. *Gene* 320: 49-58.

Craft, J. A., J. A. Gilbert, B. Temperton, K. E. Dempsey, K. Ashelford, B. Tiwari, T. H. Hutchinson, and J. K. Chipman. 2010. Pyrosequencing of *Mytilus galloprovincialis*cDNAs: Tissue-Specific Expression Patterns. *PLoS ONE* 5: e8875 EP -.

Crawford, L. 1983. The 53,000-dalton cellular protein and its role in transformation. *Int Rev Exp Pathol.* 25: 1-50.

Crooks, J. A., and H. S. Khim. 1999. Architectural vs. biological effects of a habitat-altering, exotic mussel, *Musculista senhousia*. *J. Exp. Mar. Biol. Ecol.* 240: 53-75.

Császár, N. B. M., F. O. Seneca, and M. J. H. v. Oppen. 2009. Variation in antioxidant gene expression in the scleractinian coral *Acropora millepora* under laboratory thermal stress. *Mar Ecol Prog Ser* 392: 93-102.

Cullen, P. 1990. Biomonitoring and environmental management. *Environ Monit Assess.* 14: 107-114.

da Silva, P. M., P. Soudant, M. J. Carballal, C. Lambert, and A. Villalba. 2005. Flow cytometric DNA content analysis of neoplastic cells in haemolymph of the cockle *Cerastoderma edule*. *Dis Aquat Organ* 67: 133-9.

Daly, M. J., J. D. Rioux, S. F. Schaffner, T. J. Hudson, and E. S. Lander. 2001. High-resolution haplotype structure in the human genome. *Nat Genet.* 29: 229-32.

De Laurenzi, V., Costanzo, A, Barcaroli, D, Terrinoni, A, Falco, M, Annicchiarico-Petruzzelli, M, Levrero, M, Melino, G. 1998. Two new p73 splice variants, gamma and delta, with different transcriptional activity. *J Exp Med.* 188: 1763-8.

Delaporte, M., S. Synard, J. Pariseau, P. McKenna, R. Tremblay, J. Davidson, and F. Berthe. 2008. Assessment of haemic neoplasia in different soft shell clam *Mya arenaria* populations from eastern Canada by flow cytometry. *J Invertebr Pathol* 98: 190–197.

DelMazo, J. 2006. Genetic markers and susceptibility to the effects of endocrine disruptors during mammalian testis development. *The GENDISRUPT Project*, Madrid.

Depledge, M. H., and Z. Billinghurst. 1999. Ecological Significance of Endocrine Disruption in Marine Invertebrates. *Mar Pollut Bull* 39: 32-38.

deZwart, D., W. Slooff, and J. Notenboom. 1995. Monitoring water quality in thefuture : Biomonitoring. European Union.

Dondero, F., A. Dagnino, H. Jonsson, F. CaprÃ¬, L. Gastaldi, and A. Viarengo. 2006a. Assessing the occurrence of a stress syndrome in mussels (Mytilus edulis) using a combined biomarker/gene expression approach. *Aqua Tox*

The Stavanger WorkshopBiological Effects of Environmental Pollution (BEEP) in marine coastal ecosystem The Stavanger mesocosm exposure studies 78, Supplement: S13-S24.

Dondero, F., L. Piacentini, F. Marsano, M. Rebelo, L. Vergani, P. Venier, and A. Viarengo. 2006b. Gene transcription profiling in pollutant exposed mussels (*Mytilus* spp.) using a new low-density oligonucleotide microarray. *Gene* 376: 24-36.

Dyrynda, E. A., R. K. Pipe, and N. A. Ratcliffe. 1997. Sub-populations of haemocytes in the adult and developing marine mussel. *Cell Tissue Res.* 289: 527-536.

Elston, R. A., A. S. Drum, and J. Allen, SK. 1990. Progressive development of circulating polyploid cells in *Mytilus* with haemic neoplasia. *Dis. Aquat. Org.* 8: 51-59.

Elston, R. A., M. L. Kent, and A. S. Drum. 1988. Progression, lethality and remission of hemic neoplasia in the bay mussel *Mytilus edulis*. *Dis. Aquat. Org.* 4: 135-142.

Environment_Canada. 2012. Canadian Aquatic Biomonitoring Network (CABIN).

EPD. 1993. Guidelines for Environmental Monitoring at Municipal Solid Waste Landfills. *Environmental protection*. Ministry of Environment, BC, Canada.

Ericksen, T. H., A.P. Dufour. 1986. Methods to identify water pathogens and indicator organisms. Pages 195-214. *Waterborne Diseases in the United States*. CRC Press, Boca Raton, FL.

Falconer, I., Chapman, HF, Moore, MR, Ranmuthugala, G. 2006. Endocrine-disrupting compounds: a review of their challenge to sustainable and safe water supply and water reuse. *Environ Toxicol.* 21: 181-91.

FAO, U. 1994. Integrated management of coastal zones. *FAO Fisheries Technical Paper*.

Farley, C. A. 1969. Probable neoplastic disease of the hematopoietic system in oysters, *Crassostrea virginica* and *Crassostrea gigas*. *National Cancer Institute Monographs* 31: 541-555.

Farley, C. A.. 1976. Proliferative disorders in bivalve molluscs. *Mar. Fish. Rev.* 38: 30-33.

Farley, C. A., D. L. Plutschak, and R. F. Scott. 1991. Epizootiology and distribution of transmissible sarcoma in Maryland softshell clams, *Mya arenaria,* 1984-1988. *Environ. Health Perspect.* 90: 35-41.

Feder, M., Hofmann, GE. 1999. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu. Rev. Physiol.* 61: 243-82.

Fernandes, A., Atchley WR. 2008. Biochemical and functional evidence of p53 homology is inconsistent with molecular phylogenetics for distant sequences. *J Mol Evol* 67: 51-67.

Fields, P., Zuzow, MJ, Tomanek, L. 2012. Proteomic responses of blue mussel (*Mytilus*) congeners to temperature acclimation. *J Exp Biol.* 215: 1106-16.

Fillippovich, I., Sorokina, N., Gatei, M., Haupt, Y., Hobson, K., Moallem, E., Spring, K., Mould, M., McGuckin, M. A., Lavin, M. F. and Khanna, K. K. 2001. Transactivation-deficient p73 alpha (p73 Delta exon2) inhibits apoptosis and competes with p53. *Oncogene* 20: 514-522.

Fishbein, L. 1980. Potential carcinogenic and mutagenic industrial chemicals. I. Alkylating agents. *J Toxicol Environ Health.* 6: 1133-77.

Fisher, D., D. Burton, L. Yonkos, S. Turley, and G. Ziegler. 1999. The relative acute toxicity of continuous and intermittent exposures of chlorine and bromine to aquatic organisms in the presence and absence of ammonia. *Water Res* 33.

Fossi, M. C., S. Casini, C. Savelli, C. Corbelli, E. Franchi, N. Mattei, J. C. Sanchez-Hernandez, I. Corsi, S. Bamber, and M. H. Depledge. 2000. Biomarker responses at different levels of biological organisation in crabs (*Carcinus aestuarii*) experimentally exposed to benzo(Î±)pyrene. *Chemosphere* 40: 861-874.

Franchini, A., P. Fontanili, and E. Ottaviani. 1995. Invertebrate immunocytes: relationship between phagocytosis and nitric oxide production. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 110: 403-407.

Fraser, D. K. 2011. Latency period of radiation-induced cancer. *CMAJ* 183.

Frechette', M., C. A. Butman, and W. R. Geyer. 1989. The importance of boundary-layer flows in supplying phytoplankton to the benthic suspension feeder, *Mytilus edulis* L. *Limnol. Oceanogr.* 34: 19-36.

Funes, V., J. Alhama, J. I. Navas, J. LÃ³pez-Barea, and J. Peinado. 2006. Ecotoxicological effects of metal pollution in two mollusc species from the Spanish South Atlantic littoral. *Environ Pollut* 139: 214-223.

Gagné, F., T. Burgeot, J. Hellou, S. St-Jean, E. Farcy, and C. Blaise. 2007. Spatial variations in biomarkers of *Mytilus edulis* mussels at four polluted regions spanning the Northern Hemisphere. *Environ Res.* 107: 201-17.

Gagnon, M. M., and D. A. Holdway. 2002. EROD activity, serum SDH and PAH biliary metabolites in sand flathead (*Platycephalus bassensis*) collected in Port Phillip Bay, Australia. *Mar Pollut Bul* 44: 230-237.

Gao, R., Stock, AM. 2010. Molecular strategies for phosphorylation-mediated regulation of response regulator activity. *Curr Opin Microbiol.* 13: 160-7.

Georgopoulos, P., Sasso, AF, Isukapalli, SS, Lioy, PJ, Vallero, DA, Okino, M, Reiter, L. 2009. Reconstructing population exposures to environmental chemicals from biomarkers: challenges and opportunities. *J Expo Sci Environ Epidemiol.* 19: 149-71.

Giaccia, A., and M. Kastan. 1998. The complexity of p53 modulation: emerging patterns from divergent signals. *Genes Dev.* 12: 2973-83.

Giassia,A. 2006. The complexity of p53 modulation: Emerging patterns from divergent signals. *Genes Dev.* 12: 2973-2983.

Gil, F., and A. Pla. 2001. Biomarkers as Biological Indicators of Xenobiotic Exposure. *J. Appl. Toxicol.* 21: 245–255.

Goldberg, E. D. 1975. The mussel watch - a first step in global marine monitoring. *Mar Pollut Bull* 6: 111.

Gonzalez-Tizon, A., A. Martinez-Lage, J. Ausio, and J. Mendez. 2000. Polyploidy in a natural population of mussel, *Mytilus trossulus*. *Genome* 43.

Gosling, E., ed. 1992. *Systematics and geographic distribution of Mytilus*. Elsevier Science Publishers, Amsterdam.

Gosling,E., 2003. *Bivalve Mollusks. Biology, Ecology and Culture*. Wiley-Blackwell, Hoboken, NJ.

Gracey, A. Y. 2007. Interpreting physiological responses to environmental change through gene expression profiling. *J Exp Biol* 210: 1584-1592.

Greenblatt, M., W. Bennett, M. Hollstein, and C. Harris. 1994. Mutations in the p53 tumor suppressor gene: Clues to cancer etiology and molecular pathogenesis. *Cancer Res.* 54: 4855-4878.

Gutierrez, J. L., C. G. Jones, D. L. Strayer, and O. O. Iribarne. 2003. Mollusks as ecosystem engineers: the role of shell production in aquatic habitats. *Oikos* 101: 79-90.

Health_Canada. 2010. Guidelines for Canadian Drinking Water Quality; http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/2010-sum_guide-res_recom/index-eng.php#a14.

Heath, D. D., P. D. Rawson, and T. J. Hilbish. 1995. PCR-based nuclear markers identify alien blue mussel (*Mytilus* spp.) genotypes on the west coast of Canada. *Can. J. Fish. Aquat. Sci.* 52: 2621-2627.

Higashi, R. M. 2000. Environmental Chemistry of Bioavailability in Sediments and Water Column. University of California - Davis, US EPA.

Hillman, R. 1993. Relationship of environmental contaminants to occurence of neoplasia in *Mytilus edulis* populations from east and west coast mussel-watch sites (Abstract). Pages 109. *Milford Aquaculture Seminar, 13th AGM*, Milford, Connecticut.

Hine, P. M. 1999. The inter-relationships of bivalve haemocytes. *Fish Shellfish Immunol.* 9: 367-385.

Hoeijmakers, J. H. J. 2009. DNA Damage, Aging, and Cancer. *N Engl J Med* 361: 1475-1485.

Holt, E. A., and S. W. Miller. 2011. Bioindicators: Using Organisms to Measure Environmental Impacts. *Nature Education Knowledge* 2: 8.

House, M. L., C. H. Kim, and P. W. Reno. 1998. Soft shell clams *Mya arenaria* with disseminated neoplasia demonstrate reverse transcriptase activity. *Dis. Aquat. Org.* 34: 187-192.

Innes, D., R. Thompson, and C. Couturier. 2005. A Mussel Hybrid Zone: Genetics, Ecology and Evolution. Implications for Aquaculture. Memorial University of Newfoundland.

Jayaraj, S. S., R. Thiagarajan, and S. Vincent. 2009. Encapsulation Response from the Hemocytes of Marine Mussel Perna viridis. *GJBBR* 4: 138-142.

Jessen-Eller, K., J. A. Kreiling, G. S. Begley, M. E. Steele, C. W. Walker, R. E. Stephens, and C. L. Reinisch. 2002. A new invertebrate member of the p53 gene family is developmentally expressed and responds to polychlorinated biphenyls. *Environ. Health Perspect.* 110: 377-85.

Joerger, A., and A. Fersht. 2007. Structure–function–rescue: the diverse nature of common p53 cancer mutants. *Oncogene* 26: 2226-2242.

Joerger, A.. 2010. The Tumor Suppressor p53: From Structures to Drug Discovery. *Cold Spring Harb Perspect Biol* 4.

Kaghad, M., H. Bonnet, A. Yang, L. Creancier, J. C. Biscan, A. Valent, A. Minty, P. Chalon, J. M. Lelias, X. Dumont, P. Ferrara, F. McKeon, and D. Caput. 1997. Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers. *Cell* 90: 809-19.

Kanno, T., K. Nakamura, H. Ikai, K. Kikuchi, K. Sasaki, and Y. Niwano. 2011. Literature review of the role of hydroxyl radicals in chemically-induced mutagenicity and carcinogenicity for the risk assessment of a disinfection system utilizing photolysis of hydrogen peroxide. *J Toxicol Sci.*

Kasai, H., Chung, MH, Jones, DS, Inoue, H, Ishikawa, H, Kamiya, H, Ohtsuka, E, Nishimura, S. 1991. 8-Hydroxyguanine, a DNA adduct formed by oxygen radicals: its implication on oxygen radical-involved mutagenesis/carcinogenesis. *J Toxicol Sci.* 1: 95-105.

Kato, S., S.-Y. Han, W. Liu, H. S. K. Otsuka, R. Kanamaru, and C. Ishioka. 2003. Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *PNAS* 100: 8424-8429.

Kelley, M. L., C. W. Walker, and R. J. V. Beneden. 2008. Accession number FJ041332, Direct Submission to Genbank, Submitted (13-AUG-2008) School of Marine Sciences, University of Maine, 5751 Murray Hall, Orono, ME 04469-5751, USA.

Kelley, M. L., P. Winge, J. D. Heaney, R. E. Stephens, J. H. Farell, R. J. V. Beneden, C. L. Reinisch, M. P. Lesser, and C. W. Walker. 2001. Expression of homologues for p53 and p73 in the softshell clam (*Mya arenaria*), a naturally-occurring model for human cancer. *Oncogene* 20: 748-58.

Kikuchi, S., Perrier, ND, Ituarte, P, Siperstein, AE, Duh, QY, Clark, OH. 2004. Latency period of thyroid neoplasia after radiation exposure. *Ann Surg.* 239: 536-43.

Korrida, A., S. Jadallah, H. Izaabel[1], and S. Benhissoune. 2010. Genetic Diversity Analysis of the Natural Populations of Mediterranean Mussels *[Mytilus galloprovincialis* (Lmk.)] in Agadir Bay: Assessment of the Molecular Polymorphism and Environmental Impact. *J Biol* 1: 18-25.

Kouidou, S., T. Agidou, A. Kyrkou, A. Andreou, T. Katopodi, E. Georgioua, D. Krikelis, A. Dimitriadou, P. Spanos, C. Tsilikas, H. Destouni, and G. Tzimagiorgisa. 2005. Non-CpG cytosine methylation of p53 exon 5 in non-small cell lung carcinoma. *Lung Cancer* 50: 299-307.

Kouidou, S., A. Malousi, and N. Maglaveras. 2006. Methylation and repeats in silent and nonsense mutations of p53. *Mut. Res.* 599: 167-177.

Kreiling, J. A., R. E. Stephens, and C. L. Reinisch. 2005. A mixture of environmental contaminants increases cAMP-dependent protein kinase in *Spisula* embryos. *Environ Toxicol Pharm* 19: 9-18.

Krishnakumar, P. K., E. Casillas, R. G. Snider, A. N. Kagley, and U. Varanasi. 1999. Environmental contaminants and the prevalence of hemic neoplasia (leukemia) in the common mussel (*Mytilus edulis* complex) from Puget Sound, Washington, U.S.A. *J Invertebr Pathol* 73: 135-46.

La, D., Swenberg, JA. 1996. DNA adducts: biological markers of exposure and potential applications to risk assessment. *Mutat Res.* 365: 129-46.

Lam, P. K. S. 2009. Use of biomarkers in environmental monitoring. *Ocean & Coastal Management* 52: 348-354.

Lane, D. P. 1992. p53, guardian of the genome. *Nature* 358: 15-16.

Lane, D. P., and L. V. Crawford. 1979. T antigen is bound to a host protein in SV40-transformed cells. *Nature* 278: 261-263.

Lant, B., and K. B. Storey. 2010. An Overview of Stress Response and Hypometabolic Strategies in Caenorhabditis elegans: Conserved and Contrasting Signals with the Mammalian System. *Int. J. Biol. Sci.* 6: 9-50.

Latchman, D. 2002. *Gene Regulation*. Nelson Thorns Ltd, Cheltenham, U.K.

Leavitt, D. F., J. McDowell Capuzzo, R. M. Smolowitz, D. L. Miosky, B. A. Lancaster, and C. L. Reinisch. 1990. Hematopoietic neoplasia in *Mya arenaria*;: Prevalence and indices of physiological condition. *Mar Biol.* 105: 313-321.

Liang, S., and M. Clarke. 2001. Regulation of p53 localization. *Eur. J. Biochem.* 268: 2779±2783.

López-Maury, L., S. Marguerat, and J. Bähler. 2008. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature Reviews Genetics* 9: 583-593.

Lowe, D. M., and M. N. Moore. 1978. Cytology and quantitative cytochemistry of a poliferative atypical hemocytic condition in *Mytilus edulis* (Bivalvia, mollusca). *J. Natl. Cancer Inst.* 60: 1455-9.

Lu, W.-J., J. F. Amatruda, and J. M. Abrams. 2009. p53 ancestry: gazing through an evolutionary lens. *Nat. Rev. Cancer* 9: 758-762.

Machella, N., M. Battino, B. Pisanelli, and F. Regoli. 2006. Influence of the SCGE Protocol on the Amount of Basal DNA Damage Detected in the Mediterranean Mussel, *Mytilus galloprovincialis*. *Environ Mol Mutagen* 47: 579-86.

McGladdery, S. E., C. L. Reinisch, G. S. MacCallum, R. E. Stephens, C. L. Walker, and J. T. Davidson. 2001. Haemic neoplasia in soft-shell clams (*Mya arenaria)*: Recent outbreaks in Atlantic Canada and discovery of a p53 gene homologue associated with the condition. *Bull. Aquatic Assoc. Canada* 101: 19-26.

McGrorty, S., and J. D. Goss-Custard. 1993. Population Dynamics of the *Mussel Mytilus* edulis along Environmental Gradients: Spatial Variations in Density-Dependent Mortalities. *J Anim Ecol* 62: 415-427.

Moore, J. D. 1993. Pathogenesis of Disseminated Neoplasia in Eastern Pacific mussel *Mytilus trossulus*. University of Washington, Seattle.

Moore, J. D., R. A. Elston, A. S. Drum, and M. T. Wilkinson. 1991. Alternate pathogenesis of systemic neoplasia in the bivalve mollusc *Mytilus*. *J. Invertebr. Pathol.* 58: 231-43.

Moore, M., J. Ng, and R. Lewis. 1998. Environmental poisoning: Presentation and management. *Ther Drug Monit J* 20: 502-509.

Moore, M. N. 1985. Cellular responses to pollutants. *Marine Pollution Bulletin* 16: 134-139.

Mosquin, P., Licata, AC, Liu, B, Sumner, SC, Okino, MS. 2009. Reconstructing exposures from small samples using physiologically based pharmacokinetic models and multiple biomarkers. *J Expo Sci Environ Epidemiol.* 19: 284-97.

Muttray, A., T. O'Toole, W. Morrill, R. V. Beneden, and S. Baldwin. 2010. An invertebrate mdm homolog interacts with p53 and is differentially expressed together with p53 and ras in neoplastic *Mytilus trossulus* haemocytes. *Comp Biochem Physiol B Biochem Mol Biol.* 156: 298-308.

Muttray, A., C. Reinisch, J. Miller, W. Ernst, P. Gillis, M. Losier, and J. Sherry. 2011. Haemocytic Leukemia in Prince Edward Island (PEI) soft shell clam (*Mya arenaria*): 2 spatial distribution in agriculturally impacted estuaries.

Muttray, A. 2012. Haemocytic leukemia in Prince Edward Island (PEI) soft shell clam *(Mya arenaria*): Spatial distribution in agriculturally impacted estuaries. *Sci Total Environ* 424: 130-142.

Muttray, A. F., R. L. Cox, S. D. St-Jean, P. v. Poppelen, C. L. Reinisch, and S. A. Baldwin. 2005. Identification and phylogenetic comparison of p53 in two distinct mussel species (*Mytilus*). *Comp. Biochem. Physiol. C, Toxicol Pharmacol.* 140: 237-250.

Muttray, A. 2007. Identification of DeltaN isoform and polyadenylation site choice variants in molluscan p63/p73-like homologues. *Mar. Biotechnol.* 9: 217-230.

Muttray, A. F., P. M. Schulte, and S. A. Baldwin. 2008. Invertebrate p53-like mRNA isoforms are differentially expressed in mussel haemic neoplasia. *Mar. Environ. Res.* 66: 412-421.

National_Research_Council. 1987. Committee on biological markers, Environmental Health Perspectives:74:3–9.

Nedelcu, A. M., and C. Tan. 2007. Early diversification and complex evolutionary history of the p53 tumor suppressor gene family. *Dev. Genes. Evol.* 217: 801–806.

Negoescu, A., P. Lorimier, F. Labat-Moleur, C. Drouet, C. Robert, C. Guillermet, C. Brambilla, and E. Brambilla. 1996. In situ apoptotic cell labeling by the TUNEL method: improvement and evaluation on cell preparations. *J Histochem Cytochem.* 44: 959-68.

NSSP. 2009. National Shellfish Sanitation Program2009 Section VIII Interpretation Number: 09-III-@.02-100, National Shellfish Sanitation Program, Guide for the Control of Molluscan Shellfish in U. S. F. a. D. Administration, S. S. Team, and D. o. C. Program, eds.

Ollmann, M., L. M. Young, C. J. Di Como, F. Karim, M. Belvin, S. Robertson, K. Whittaker, M. Demsky, W. W. Fisher, A. Buchman, G. Duyk, L. Friedman, C. Prives, and C.

Kopczynski. 2000. *Drosophila* p53 is a structural and functional homolog of the tumor suppressor p53. *Cell* 101: 91-101.
Pariseau, J., P. McKenna, M. AboElkhair, R. Saint-Louis, Ã. m. Pelletier, T. Davidson, R. j. Tremblay, F. Berthe, and A. Siah. 2011. Effects of pesticide compounds (chlorothalonil and mancozeb) and benzopyrene mixture on aryl hydrocarbon receptor, p53 and ubiquitin gene expression levels in haemocytes of soft-shell clams (*Mya arenaria*). Pages 1765-1772. *Ecotoxicology*. Springer Netherlands.
Peters, L. D., J. P. Shaw, M. Nott, S. C. M. O'Hara, and D. R. Livingstone. 1999. Development of cytochrome P450 as a biomarker of organic pollution in i sp.: field studies in United Kingdom ('Sea Empress' oil spill) and the Mediterranean Sea. *Biomarkers* 4: 425-441.
Petitjean, A., E. Mathe, S. Kato, C. Ishioka, S. Tavtigian, P. Hainaut, and M. Olivier. 2007. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: Lessons from recent developments in the IARC TP53 database. *Hum. Mutat.* 6: 622-629.
Petrocelli, S., A. Hanks, and J. Anderson. 1973. Uptake and Accumulation of an Organochlorine Insecticide (Dieldrin) by an Estuarine Mollusc, *Rangia cuneata*. *Bull Environ Contam Toxicol* 10: 315-320.
Picado, A., M. J. Bebianno, M. H. Costa, A. Ferreira, and C. Vale. 2007. Biomarkers: a strategic tool in the assessment of environmental quality of coastal waters. *Hydrobiologia* 587: 79-87.
Picard, D. J., and P. M. Schulte. 2004. Variation in gene expression in response to stress in two populations of *Fundulus heteroclitus*. *Comp Biochem and Physiol- Part A: Mol &; Integr Physiol* 137: 205-216.
Pipe, R. K. 1990. Differential binding of lectins to haemocytes of the mussel *Mytilus edulis*. *Cell Tissue Res.* 261: 261-268.
Pipe, R. K., S. R. Farley, and J. A. Coles. 1997. The separation and characterisation of haemocytes from the mussel *Mytilus edulis*. *Cell Tissue Res.* 289: 537-545.
Pisoni, M., Cogotzi, L, Frigeri, A, Corsi, I, Bonacci, S, Iacocca, A, Lancini, L, Mastrototaro, F, Focardi, S, Svelto, M. 2004. DNA adducts, benzo(a)pyrene monooxygenase activity, and lysosomal membrane stability in *Mytilus galloprovincialis* from different areas in Taranto coastal waters (Italy). *Environ Res.* 96: 163-75.
Place, S. P., B. A. Menge, and G. E. Hofmann. 2012. Transcriptome profiles link environmental variation and physiological response of *Mytilus californianus* between Pacific tides. *Funct Ecol* 26: 144–155,.
Polo, S., and S. Jackson. 2011. Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes & Dev.* 25: 409-433.
Poly, W. 1997. Nongenetic variation, genetic-environmental interactions and altered gene expression. I. Temperature, photoperiod, diet, pH and sex-related effects. *Comp Biochem Physiol A Physiol.* 117: 11-66.
Ponassi, R., A. Terrinoni, A. Chikh, A. Rufini, A. Lena, B. Sayan, G. Melino, and E. Candi. 2006. p63 and p73, members of the p53 gene family, transactivate PKCdelta. *Biochem Pharmacol* 72: 1417-22.
Porte, C., G. Janer, L. C. Lorusso, M. Ortiz-Zarragoitia, M. P. Cajaraville, M. C. Fossi, and L. Canesi. 2006. Endocrine disruptors in marine organisms: Approaches and perspectives. *Comp. Biochem. Physiol., Part C: Toxicol. Pharmacol.* 143: 303-315.
Rasmussen, L. P. D. 1986. Occurrence, prevalence and seasonality of neoplasia in the marine mussel *Mytilus edulis* from three sites in Denmark. Pages 59-64. *Marine Biology*. Springer Berlin / Heidelberg.

Ratcliffe, N., and A. Rowley. 1981. *Invertebrate blood cells. Mollusks.* Academic Press, New York.

Rawson, P. D., and T. J. Hilbish. 1995. Distribution of male and female mtDNA lineages in populations of blue mussels, *Mytilus trossulus* and *M. galloprovincialis*, along the Pacific coast of North America. *Mar Biol.* 124: 245-250.

Rawson, P. D., K. L. Joyner, K. Meetze, and T. J. Hilbish. 1996. Evidence for intragenic recombination within a novel genetic marker that distinguishes mussels in the *Mytilus* edulis species complex. *Heredity* 77: 599-607.

Rawson, P. D., S.Hayhurst, and B. Vanscoyoc. 2001. Species composition of blue mussel populations in the northeastern Gulf of Main. *J. Shellfish Res.* 20: 31-38.

Reinisch, C. L., A. M. Charles, and A. M. Stone. 1984. Epizootic neoplasia from softshell clams collected from New Bedford Harbour. *Haz Waste* 1: 73-81.

Roh, M., O. Franco, S. Hayward, R. v. d. Meer, and S. Abdulkadir. 2008. A Role for Polyploidy in the Tumorigenicity of Pim-1-Expressing Human Prostate and Mammary Epithelial Cells. *PLoS ONE* 3.

Romalde, J. L., and J. L. Barja. 2010. Bacteria in molluscs: good and bad guys in A. Mendez-Vilas, ed. *Current research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology*.

Roper, J. M., D. S. Cherry, J. W. Simmers, and H. E. Tatem. 1996. Bioaccumulation of toxicants in the zebra mussel, *Dreissena polymorpha*, at the Times Beach Confined Disposal Facility, Buffalo, New York. *Env Pollut.* 94: 117-129.

Rosenberg, D., and V. Resh. 1993. *Freshwater biomonitoring and benthic macroinvertebrates.* Chapman and Hall, New York.

Rotchell, J., Clarke, KR, Newton, LC, Bird, DJ. 2001. Hepatic metallothionein as a biomaker for metal contamination: age effects and seasonal variation in European flounders (*Pleuronectes flesus*) from the Severn Estuary and Bristol Channel. *Mar Environ Res.* 52: 151-71.

Ruaro, E., L. Collavin, G. D. Sal, R. Haffner, M. Oren, A. Levine, and C. Schneider. 1997. A proline-rich motif in p53 is required for transactivation- independent growth arrest as induced by Gas1. *PNAS* 94: 4675-4680.

Ruths, D., and L. Nakhleh. 2005. Recombination and phylogeny: effects and detection. *Int. J. Bioinformatics Research and Applications* 1: 202-212.

Rutkowski, R., K. Hofmann, and A. Gartner. 2010. Phylogeny and Function of the Invertebrate p53 Superfamily. *Cold Spring Harb Perspect Biol*.

Salazar, M., and S. Salazar. 1996. Mussels as bioindicators: effects of tbt on survival, bioaccumulation, and growth under natural conditions. Pages 305-327. Chapman & Hall, London.

Sarkar, A., D. Ray, A. Shrivastava, and S. Sarker. 2006. Molecular Biomarkers: Their significance and application in marine pollution monitoring. *Ecotoxicology* 15: 333-340.

Schmitt, C., J. Fridman, M. Yang, E. Baranov, R. Hoffman, and S. Lowe. 2001. Dissecting p53 tumor suppressor functions in vivo. *Cancer Cell* 1: 289-98.

Schulte, P. M. 2004. Changes in gene expression as biochemical adaptations to environmental change: a tribute to Peter Hochachka. *Comp Biochem Physiol B Biochem Mol Biol.* 139: 519-529.

Schumacher, B., K. Hofmann, S. Boulton, and A. Gartner. 2001. The C. elegans homolog of the p53 tumor suppressor is required for DNA damage-induced apoptosis. *Curr Biol* 11: 1722-7.

Shabalina, S., A. Ogurtsov, and N. Spiridonov. 2006. A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.* 34: 2428-2437.

Shaw, P. 1996. The role of p53 in cell cycle regulation. *Pathol Res Pract.* 192: 669-75.
Shen, L., J. Basilion, and V. S. Jr. 1999. Single-nucleotide polymorphisms can cause different structural folds of mRNA. *Proc Natl Acad Sci U S A.* 96: 7871-6.
Sherman, K. 1994. Sustainability, biomass yields, and health of coastal ecosystems: an ecological perspective. *Mar. Ecol. Prog. Ser.* 112: 277-301.
Siah, A., M. Delaporte, J. Pariseau, P. McKenna, and F. C. J. Berthe. 2008. Patterns of p53, p73 and mortalin gene expression associated with haemocyte polyploidy in the soft-shell clam, *Mya arenaria*. *J Invert Pathol* 98: 148-152.
Silins, I., and J. Hogberg. 2011. Combined Toxic Exposures and Human Health: Biomarkers of Exposure and Effect. *International Journal of Environmental Research and Public Health* 8: 629-647.
Smolarz, K., T. Renault, and M. Wolowicz. 2005. Histology, cytogenetics and cytofluorymetry in diagnosis of neoplasia in *Macoma balthica (Bivalvia, L.)* from the southern Baltic Sea. *Caryologia* 58: 212-219.
Smolowitz, R., and D. Leavitt. 1996. Neoplasia and other pollution associated lesions in *Mya arenaria* from Boston Harbor. *J.Shellfish Res.* 15: 520.
Smolowitz, R. M., D. Miosky, and C. L. Reinisch. 1989. Ontogeny of leukemic cells of the soft shell clam. *J. Invert. Pathol.* 53: 41-51.
Smolowitz, R. M., and C. L. Reinisch. 1993. A novel adhesion protein expressed by ciliated epithelium, hemocytes, and leukemia cells in soft-shell clams. *Developmental and Comparative Immunology* 17: 475-81.
Snyder, M., Girvetz, E, Mulder, EP. 2001. Induction of marine mollusc stress proteins by chemical or physical stress. *Arch Environ Contam Toxicol.* 41: 22-9.
Snyder, M., and S. Rossi. 2004. Stress protein (HSP70 family) expression in intertidal benthic organisms: the example of Anthopleura elegantissima (Cnidaria: Anthozoa). *Sci Mar* 68: 155-162.
Soto, A. M., and C. Sonnenschein. 2010. Environmental causes of cancer: endocrine disruptors as carcinogens. *Nat Rev Endocrinol.* 6: 363-370.
Soussi, T. 2012. p53 Pathways. *The p53 Web Site.*
Spoel, S. H., Tada, Y. and Loake, G. J. 2010. Post-translational protein modification as a tool for transcription reprogramming. *New Phytologist* 186: 333-339.
Stegeman, J., Brouwer, M. , Di Giulio, R.T. , Forlin, L. , Fowler, B.A. , Sanders, B.M. , Van Veld, P.A. 1992. Molecular responses to environmental contamination: enzyme and protein systems as indicators of chemical exposure and effect. Pages 253–310 in R. A. K. R.J. Huggett, P.M. Mehrle Jr., H.L. Bergman, ed. *Biomarkers: Biochemical, Physiological, and Histological Markers of Anthropogenic Stress*. Lewis Publishers, Boca Raton.
Stephens, R. E., C. W. Walker, and C. L. Reinisch. 2001. Multiple protein differences distinguish clam leukemia cells from normal hemocytes: Evidence for the involvement of p53 homologues. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* 129: 329-338.
Stifanic, M., M. Micic, A. Ramšak, S. Blaškovic, A. Ruso, R. K. Zahn, and R. Batel. 2009. p63 in *Mytilus galloprovincialis* and p53 family members in the phylum Mollusca. *Comp. Biochem. Physiol. Part B* 154: 264–273.
Stillman, J., and G. Somero. 1996. Adaptation to temperature stress and aerial exposure in congeneric species of intertidal porcelain crabs (genus *Petrolisthes*): correlation of physiology, biochemistry and morphology with vertical distribution. *J Exp Biol* 199: 1845–1855.

St-Jean, S. D., S. C. Courtenay, and R. W. Parker. 2003. Immunomodulation in blue mussels (*Mytilus edulis*) exposed to a pulp and paper mill effluent in Eastern Canada. *Water Qual Res J Can* 38: 647-666.

St-Jean, S. D., R. E. Stephens, S. C. Courtenay, and C. L. Reinisch. 2005. Detecting p53 family proteins in leukemia cells of *Mytilus edulis* from Pictou Harbour, Nova Scotia. *Can. J. Fish. Aquat. Sci.* 62: 2055-2066.

Storchova, Z., and D. Pellman. 2004. From polyploidy to aneuploidy, genome instability and cancer. *Nat Rev Mol Cell Biol.* 5: 45-54.

Storey, K. 2004. Gene regulation in physiological stress. *Intl Congress Ser* 1275: 1-13.

Strauss, B. 2000. Role in tumorigenesis of silent mutations in the TP53 gene. *Mutat Res.* 457: 93-104.

Strickland, P. T., and J. D. Groopman. 1995. Biomarkers for assessing environmental exposure to carcinogens in the diet. *Am J Clin Nutr* 61: 710S-720S.

Tegler, B., Sharp, M, Johnson, MA. 2001. Ecological Monitoring and Assessment Network's proposed core monitoring variables: an early warning of environmental change. *Environ Monit Assess.* 67: 29-56.

The_International_HapMap_Consortium. 2003. The International HapMap Project. *Nature* 426.

Thier, R., T. Bruning, P. H. Roos, H. Rihs, K. Golka, Y. Ko, and H. M. Bolt. 2003. Markers of genetic susceptibility in human environmental hygiene and toxicology: The role of selected CYP, NAT and GST genes. *Int J Hyg Envir Heal.* 206: 149-171.

Tiu, R., L. Gondek, C. O'Keefe, and J. Maciejewski. 2007. Clonality of the stem cell compartment during evolution of myelodysplastic syndromes and other bone marrow failure syndromes. *Leukemia* 21: 1648-1657.

Toledo, F., and G. M. Wahl. 2006. Regulating the p53 pathway: in vitro hypotheses, in vivo veritas. *Nat Rev Cancer* 6: 909-23.

Tornaletti, S., and G. Pfeifer. 1995. Complete and tissue-independent methylation of CpG sites in the p53 gene: implications for mutations in human cancers. *Oncogene* 10: 1493-9.

Toro, J., D. J. Innes, and R. J. Thompson. 2004. Genetic variation among life-history stages of mussels in a *Mytilus edulis–M. trossulus* hybrid zone. *Mar Biol* 145: 713-725.

Van Beneden, R. J. 1994. Molecular analysis of bivalve tumors: models for environmental/genetic interactions. *Environ Health Persp Supplements* 102: 81-83.

Van Beneden, R. J.. 1997. Environmental effects and aquatic organisms: Investigations of molecular mechanisms of carcinogenesis. *Environmental Health Perspectives Supplements* 105: 669-.

VanEgmond, H. P., VanApeldoorn, M.E., Speijers, G.J.A. 2004. Marine Biotoxins. *Natural Toxins, RIVM project*. Food And Agriculture Organization Of The United Nations, Rome.

Vassilenko, E. I., A. F. Muttray, P. M. Schulte, and S. A. Baldwin. 2010. Variations in p53-like cDNA sequence are correlated with mussel haemic neoplasia: A potential molecular-level tool for biomonitoring. *Mutat Res.*;701(2): 145-52.

Venier, P., and S. Canova. 1996. Formation of DNA adducts in the gill tissue of *Mytilus galloprovincialis* treated with benzo[a]pyrene. *Aquat Toxicol* 34: 119-133.

Venier, P., C. De Pitta, A. Pallavicini, F. Marsano, L. Varotto, C. Romualdi, F. Dondero, A. Viarengo, and G. Lanfranchi. 2006. Development of mussel mRNA profiling: Can gene expression trends reveal coastal water pollution? *Mutat. Res. Fundam. Mol. Mech. Mugag.* 602: 121-134.

Venier, P., De Pittà, C, Bernante, F, Varotto, L, De Nardi, B, Bovo, G, Roch, P, Novoa, B, Figueras, A, Pallavicini, A, Lanfranchi, G. 2009. MytiBase: a knowledgebase of mussel (*M. galloprovincialis)* transcribed sequences. *BMC Genomics.* 10.

Venot, C., M. Maratrat, C. Dureuil, E. Conseiller, L. Bracco, and L. Debussche. 1998. The requirement for the p53 proline-rich functional domain for mediation of apoptosis is correlated with specific PIG3 gene transactivation and with transcriptional repression. *EMBO J* 17: 4668–4679.

Viarengo, A., D. Lowe, C. Bolognesi, E. Fabbri, and A. Koehler. 2007. The use of biomarkers in biomonitoring: A 2-tier approach assessing the level of pollutant-induced stress syndrome in sentinel organisms. *Comp Biochem Physiol C* 146: 281–300.

Vries, A. d., E. Flores, B. Miranda, H. Hsieh, C. v. Oostrom, J. Sage, and T. Jacks. 2002. Targeted point mutations of p53 lead to dominant-negative inhibition of wild-type p53 function. *Proc Natl Acad Sci U S A* 99: 2948-53.

Wada, H., K. G. Salvante, E. Wagner, T. D. Williams, and C. W. Breuner. 2009. Ontogeny and Individual Variation in the Adrenocortical Response of Zebra Finch (*Taeniopygia guttata*) Nestlings. *Physiol Biochem Zool* 82: 325-331.

Walker, C., S. Bottger, and B. Low. 2006. Mortalin-based cytoplasmic sequestration of p53 in a nonmammalian cancer model. *Am J Pathol* 168: 1526-1530.

Walker, C., R. VanBeneden, A. Muttray, S. Böttger, M. Kelley, A. Tucker, and W. Thomas. 2011. p53 Superfamily Proteins in Marine Bivalve Cancer and Stress Biology. *Adv Mar Biol.* 59: 1-36.

Wheelan, S. J., D. M. Church, and J. M. Ostell. 2001. Spidey: a tool for mRNA-to-genomic alignments. *Genome Res.* 11: 1952-1957.

White, G. C. 1986. *Handbook of chlorination, 2nd edition*. Van Nostrand Reinhold Company Inc., New York.

Widdows, J., P. Donkin, F. J. Staff, P. Matthiessen, R. J. Law, Y. T. Allen, J. E. Thain, C. R. Allchin, and B. R. Jones. 2002. Measurement of stress effects (scope for growth) and contaminant levels in mussels (*Mytilus edulis*) collected from the Irish Sea. *Mar Environ Res.* 53: 327-356.

Wu, R., Siu, WH, Shin, PK. 2005. Induction, adaptation and recovery of biological responses: Implications for environmental monitoring. *Mar Pollut Bull.* 51: 623-34.

Yang, A., M. Kaghad, Y. Wang, E. Gillett, M. D. Fleming, V. Dotsch, N. C. Andrews, D. Caput, and F. McKeon. 1998. p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. *Molecular Cell* 2: 305-316.

Yang, A., Walker, N, Bronson, R, Kaghad, M, Oosterwegel, M, Bonnin, J, Vagner, C, Bonnet, H, Dikkes, P, Sharpe, A, McKeon, F, Caput, D. 2000. p73-deficient mice have neurological, pheromonal and inflammatory defects but lack spontaneous tumours. *Nature* 404: 99-103.

Yeom, D.-H., and S. M. Adams. 2007. Assessing effects of stress across levels of biological organization using an aquatic ecosystem health index. *Ecotox Environ Safe* 67: 286-295.

Yevich, P. P., and C. A. Barszcz. 1978. Neoplasia in soft-shell clams (*Mya arenaria*) collected from oil-impacted sites. *Ann N Y Acad Sci,* 298: 409-26.

Zaika, A. I., N. Slade, S. H. Erster, C. Sansome, T. W. Joseph, M. Pearl, E. Chalas, and U. M. Moll. 2002. DeltaNp73, a dominant-negative inhibitor of wild-type p53 and TAp73, is up-regulated in human tumors. *J. Exp. Med.* 196: 765-780.

Appendix A. Haemocyte aggregates

Assessment of mussel leukemia using cell appearance and cell ratios in mussel hemolymph is complicated by several factors, one of which is the formation of cell aggregates. Structure and role of these formations were not studied and discussed in details, therefore I decided to include my observations and hypothesis regarding the agglomerates formation. Results of the Comet assay, a technique developed for DNA integrity analysis, support my hypothesis regarding the nature of the haemocyte agglomerates.

Method: Comet assay

Hemolymph (500 µL) was centrifuged at 400xg for 10 min at 6^{o}C, supernatant was removed and cells were re-suspended in 500 µL of 0.01M PBS with 2% NaCl. To induce DNA damage, 500 µL of cell suspension was incubated with 100 µL of 30 mM H_2O_2 at 3^{o}C for 15 min, then re-suspended in 500 µL of 0.01M PBS with 2% NaCl . For each sample, 50 µL of cell suspension was mixed with 500 µL of warm liquid low melting point agarose (LMP) and applied on Trevigen Comet Assay® HT slides (20 Well 4252-200-01), 30 µL per spot. Slides with solidified agarose were immersed in the lysis solution (0.03M NaOH, 1.2M NaCl and 0.1% N-Lauroylsarcosine, according to the protocols kindly provided by P. Olive (personal communications) for 1h at 3^{o}C. After lysis, slides were placed into the freshly prepared alkaline solution (0.03M NaOH, 2mM disodium EDTA, P. Olive protocols, personal communications) for another 1hour. Alkaline electrophoresis was done in 300mM NaOH, 1mM EDTA buffer at 20V for 20min. Slides were washed 2 times in dH_2O for 10 min, fixed in 70% ethanol for 5 min and air dried. For staining, 100 µL of diluted SYBR Green I were added to each spot for 30 min, excess stain was removed, slides were allowed to dry completely in the dark and were analyzed by epifluorescence microscopy with the appropriate filters (SYBR Green has excitation and emission wavelengths 495nm and 521nm, respectively).

Results

Based on the available information, discussed in the Introduction in section 1.2.3 “Mussel Haematocytology” and our observations, it was suggested that at least some cell aggregates are formed during the host cell response to the presence of foreign particles and

classified them based on the visual appearance. Thus, Figure A1(b) shows a typical example of an aggregate of cells that was seen in mussel hemolymph, which were classified as the early stage of *phagocytosis*, adhesion, when cells without pseudopodia, most likely neoplastic cells, are attached to the surface of the normal haemocyte. Figure A1 (c,d) show another type of the host response, the encapsulation process Figure A1 (c) shows normal haemocytes in concentric layers around the group of neoplastic-looking cells, too large to be phagocytosed. Figure A1 (d) shows capsule-like structures, surrounded by the normal haemocytes. Figure A1 (e,f) shows late, sub-lethal stage of the leukemia with the overwhelming amount of neoplastic cells. It was tentatively suggested that the darker areas on the panel (e) may be the former capsule-like structures containing partially degraded neoplastic cells.

To further investigate if these cell aggregates contained cells, a DNA integrity analysis or Comet assay was performed and the results were compared to ones for normal un-aggregated granulocytes. Figure A2 (a), shows normal haemocytes with intact DNA, that does not move during a short electrophoresis used in this assay. When haemocyte contains fragmented DNA, the smaller parts of it will move and form "tails" of the size and density proportional to the damage rate. Haemocytes with damaged DNA create comet-like shapes presented on Figure A2 (b). Capsule-like aggregates or aggregates, form huge "comets" presented on the Figure A1(c), with a long and dense "tails", which indicates the presence of large number of partially degraded DNA, most likely originating from the partially digested cells within the agglomerate.

Our observations of the cell morphology and capsule-like structures in combination with the results of the comet assay suggest that some cell aggregates in the hemolymph could consist of neoplastic cells surrounded by normal granulocytes as a sign of remission or an attempt at elimination of the neoplastic cells from the hemolymph. This process was briefly discussed by Elston et al 1988 and never been confirmed or disproved. The data provided in my study is very preliminary. As well, there may be other reasons for mussel haemocytes to aggregate. Nevertheless, the investigation of the neoplasia remission concept will significantly improve our understanding of the disease and its use as an endpoint for biomonitoring.

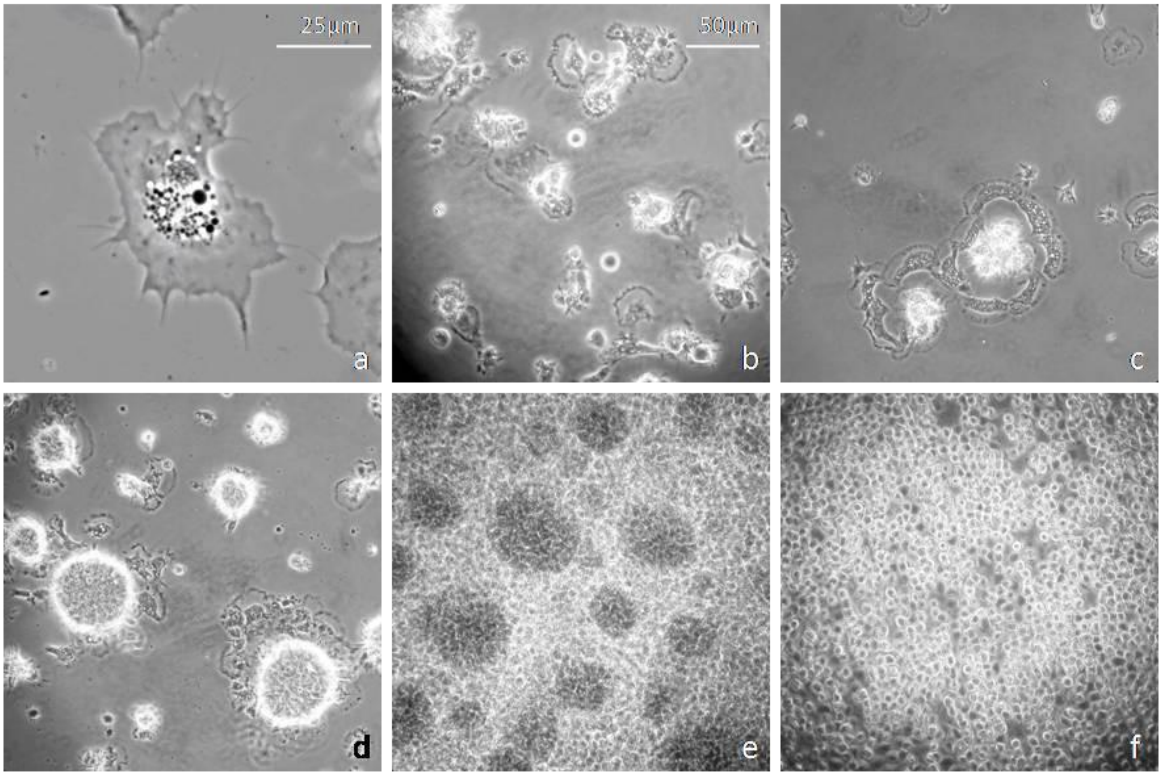


Figure A 1. Haemocyte types and cell aggregates. a: normal granulocyte; b: neoplastic cells attached to the surface of normal granulocyte; c: concentric layer of normal granulocyte around aggregate of neoplastic cells; d: capsule-like structure. e,f: sublethal stage of the leukemia. Panels b-f are in the same scale.

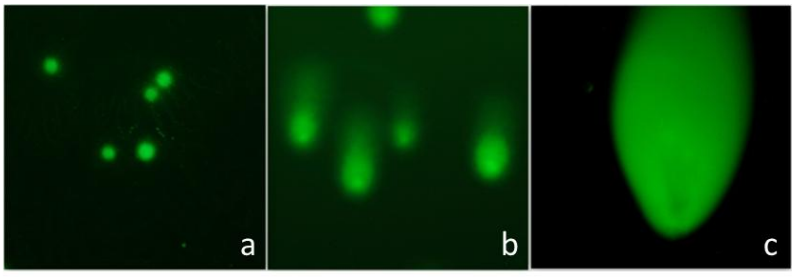


Figure A 2. DNA integrity analysis. analysis (comet assay). a: individual normal cells with intact DNA, (40x); b: cells after the exposure to DNA damaging agent (40x); c: capsule-like cell aggregate (10x).

Appendix B. Analysis of *M. trossulus* p53 variations using HRM-PCR

In order to develop a more cost and time effective and non-sequencing way to detect specific SNPs a High Resolution Melting (HRM)–qPCR technique was used to detect leukemia associated SNPs in *M. trossulus* p53.

M. trossulus p53 cDNA fragment was amplified by PCR as described in Chapter 2 (Materials and Methods). Roche Light Cycler 480 HRM Master mix (2x concentration) was used according to the product instructions for 20uL reactions on plates Applied Biosystems MicroAmp Fast Optical (96 wells). All reactions were performed in triplicates on LightCycler 480 Real-Time PCR System. Primer pair F773 (ACCTCAAGCTGGCTCAGAAT) -R856 (GGTCTTCTGTTTGGTCCTCCT) allowed to distinguish genotypes at position 821 bp. Another primer pair F334 (CAATCACATCACCTCCACCT) R432 (TGAAGCCATAATCTCCTGGAT) allowed to distinguish genotypes at position 392bp.

This method might be a good alternative for Sanger sequencing for *M. trossulus* genotyping since the method is more time efficient. Although, there is a significant limitation of the method, it would allow the recognizing of only one variable position per analyzed fragment. In situations when there are more than one SNPs in the amplified fragment the results might be misleading. It was suggested that the method should not be used when mussel population might include interspecies hybrids since hybridization might cause high variability in the gene sequence as it was discussed in Chapter 3. Hybridization and subsequent recombination might increase variability of p53 sequence as it was detected for another mussel species, *M. edulis* (Chapter 3).

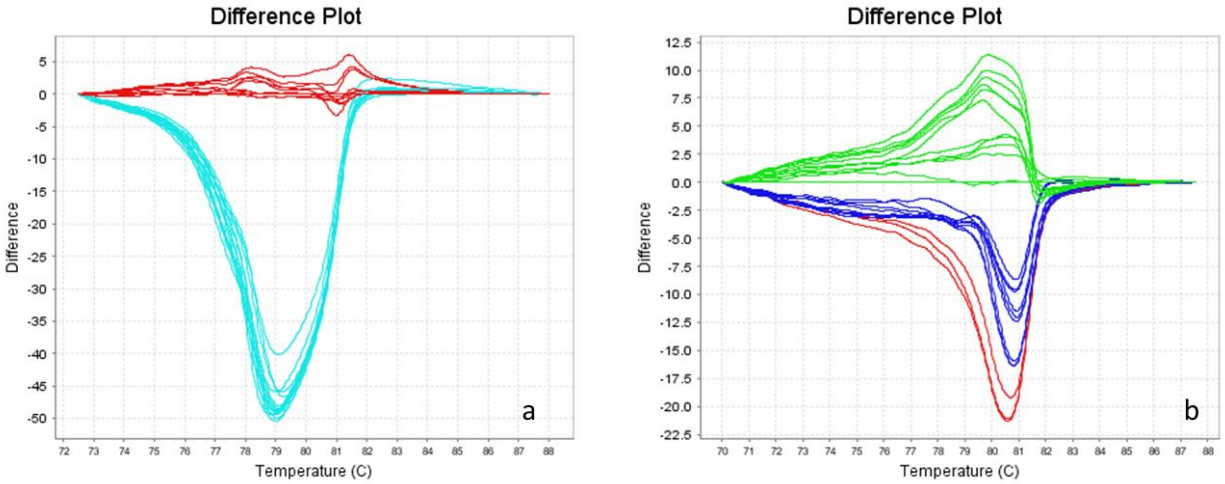


Figure A 3. Detection of SNPs in *M. trossulus* p53 coding region using High Resolution Melting Analysis (HRM-qPCR). a: SNP at position 392bp; red -variant GT, blue- variant TT. b: SNP at position 821bp; red and blue -variant CC, green- variant CT.

Appendix C. *M. edulis* and *M. trossulus* genotypes and theoretical haplotypes.

Table A 1. *M. edulis* genotypes. The first column (L%) shows the percentage of neoplastic-looking cells in the hemolymph sample. The second column indicates the phenotype groups defined in Table 4.2: L-late leukemic, shaded, Tr – transitional, N- normal.

L%	phen	snp158	snp182	snp191	snp204	snp219	snp231	snp261	snp262	snp267	snp297	snp347	snp380	snp386	snp392	snp401	snp407	snp416	snp419	snp428	snp465	snp524	snp539	snp543	snp545	snp552	snp560	snp563	snp566	snp575	snp605	snp680	snp731	snp737	snp740	snp767	snp821	snp854	snp914	snp917
M. trossulus																																								
0	N	TT	TT	TT	AA	CC	CC	CC	TT	GG	CC	CC	GG	AA	GT	AA	TT	AA	AA	CC	AA	AA	AA	AA	AA	GG	CC	AA	GG	AA	TT	CC	CC	CC	AA	TT	CC	AA	TT	TT
1	N	TT	TT	TT	AA	CC	CC	CC	TT	GG	CC	CC	GG	AA	GT	AA	TT	AA	AA	CC	AA	AA	AA	AA	AA	GG	CC	AA	GG	AA	TT	CC	CC	CC	AA	TT	CC	AA	TT	TT
0	N	TT	TT	TT	AA	CC	CC	CC	TT	GG	CC	CC	GG	AA	GT	AA	TT	AA	AA	CC	AA	AA	AA	AA	AA	GG	CC	AA	GG	AA	TT	CC	CC	CC	AA	TT	CC	AA	TT	TT
0	N	TT	TT	TT	AA	CC	CC	CC	TT	GG	CC	CC	GG	AA	GT	AA	TT	AA	AA	CC	AA	AA	AA	AA	AA	GG	CC	AA	GG	AA	TT	CC	CC	CC	AA	TT	CC	AA	TT	TT
0	N	TT	TT	TT	AA	CC	CC	CC	TT	GG	CC	CC	GG	AA	GT	AA	TT	AA	AA	CC	AA	AA	AA	AA	AA	GG	CC	AA	GG	AA	TT	CC	CC	CC	AA	TT	CC	AA	TT	TT
100	L	TT	CT	TT	AA	CC	CC	CC	TT	GG	CC	CC	GG	AA	GG	AA	TT	AA	AA	CC	AA	AA	AA	AA	AA	GG	CC	AA	GG	AA	TT	CC	CC	CC	AA	TT	CT	AA	TT	TT
100	L	TT	CT	TT	AA	CC	CC	CC	TT	GG	CC	CC	GG	AA	GG	AA	TT	AA	AA	CC	AA	AA	AA	AA	AA	GG	CC	AA	GG	AA	TT	CC	CC	CC	AA	TT	CT	AA	TT	TT
100	L	TT	CT	TT	AA	CC	CC	CC	TT	GG	CC	CC	GG	AA	GG	AA	TT	AA	AA	CC	AA	AA	AA	AA	AA	GG	CC	AA	GG	AA	TT	CC	CC	CC	AA	TT	CT	AA	TT	TT
100	L	TT	CT	TT	AA	CC	CC	CC	TT	GG	CC	CC	GG	AA	GG	AA	TT	AA	AA	CC	AA	AA	AA	AA	AA	GG	CC	AA	GG	AA	TT	CC	CC	CC	AA	TT	CT	AA	TT	TT
100	L	TT	CT	TT	AA	CC	CC	CC	TT	GG	CC	CC	GG	AA	GG	AA	TT	AA	AA	CC	AA	AA	AA	AA	AA	GG	CC	AA	GG	AA	TT	CC	CC	CC	AA	TT	CT	AA	TT	TT
M. edulis																																								
100	L	TC	TC	TT	AA	CC	CC	CC	TT	GG	CC	TT	GG	AA	GG	AA	TT	AA	AA	CC	AA	AA	AA	TT	AA	GG	CC	AA	GG	AA	TT	CC	CC	CC	CA	CC	TC	AA	GG	AA
100	L	TC	TC	TT	GA	TC	CC	CC	TT	GG	TC	TC	GA	CA	TG	AA	AT	AA	AA	GC	CA	AA	AA	TT	AA	GA	CC	AA	GG	AT	TT	TC	TC	TC	CA	CC	TC	GA	CC	CC
100	L	TC	TC	TT	AA	CC	CC	CC	TT	GG	CC	TT	GG	AA	GG	AA	TT	AA	AA	CC	AA	AA	AA	TT	AA	GG	CC	AA	GG	AA	TT	CC	CC	CC	AA	CC	TC	AA	GG	AA
100	L	TC	TC	TT	AA	CC	CC	CC	TT	GG	CC	TT	GG	AA	GG	AA	TT	AA	AA	CC	AA	AA	AA	TT	AA	GG	CC	AA	GG	AA	TT	CC	CC	CC	AA	TT	TC	AA	GG	AA
100	L	TC	TC	TT	AA	CC	CC	CC	TT	GG	CC	TT	GG	AA	GG	AA	TT	AA	AA	CC	AA	AA	AA	TT	AA	GG	CC	AA	GG	AA	TT	CC	CC	CC	AA	CC	TC	AA	AA	AA
100	L	TC	TC	TT	AA	CC	CC	CC	TT	GG	CC	TT	GG	AA	GG	AA	TT	AA	AA	CC	AA	AA	AA	TT	AA	GG	CC	AA	GG	AA	TT	CC	CC	CC	AA	CC	TC	AA	CC	CC
100	L	TC	TC	TT	GA	TC	CC	TC	AT	TG	TC	TC	GA	AA	TG	AA	AT	CA	CA	GC	CA	AA	AA	TT	AA	GG	CC	AA	GG	AT	TT	TC	TC	TC	CA	CC	TC	GA	CC	CC
76	Tr	CC	TT	TT	GA	TT	CC	TT	AA	TT	TT	TC	GA	CA	TT	AA	AA	CC	CC	GG	CA	AA	AA	TT	AA	GG	CC	AA	GG	TT	TT	TT	TT	TT	CC	CC	CC	GG	CC	CC
60	Tr	CC	TT	TT	GG	TT	CC	TT	AT	TT	TT	CC	GA	CA	TT	AA	AA	CC	CC	GG	CC	AA	AA	TT	AA	GA	CC	AA	GG	TT	TT	TT	TT	TC	CC	TT	CC	GG	AA	AA
50	Tr	CC	TT	TT	GA	TT	TC	TT	AA	TT	TT	CC	AA	AA	TT	AA	AA	CC	CC	GG	CA	AT	GA	TT	GA	GG	TC	AA	GA	TT	TT	TT	TT	TT	CC	TT	CC	GG	AA	AA
46	Tr	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	GA	CA	TT	AA	AA	CC	CC	GG	AA	AA	GA	TT	GA	AA	CC	AA	GG	TT	TC	TT	TT	TT	CC	CC	TC	GG	GG	AA
41	Tr	CC	TT	AT	GA	TT	CC	TT	AA	TT	TC	TT	GG	AA	TG	AA	AT	AA	AA	CC	AA	AA	AA	TT	AA	GG	CC	AA	GG	AA	TT	TC	CC	CC	AA	CC	TC	AA	CC	CC
40	Tr	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	AA	TT	AA	AA	CC	CC	GG	AA	TT	GG	TT	GG	GG	TT	AA	AA	TT	TT	TT	TT	TT	CC	TT	CC	GG	GG	AA
40	Tr	CC	TT	AT	GG	TT	CC	TT	AA	TT	TT	CC	GA	CA	TT	AA	AA	CC	CC	GG	CA	AT	GA	TT	GA	GA	TC	AA	GA	TT	TT	TC	TT	TT	CC	TC	CC	AA	AA	AA
40	Tr	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CA	TT	AA	AA	CC	CC	GG	CC	AA	AA	TT	AA	GA	CC	AA	GG	TT	TT	TT	TT	TT	CC	CC	CC	GG	CC	CC
33	Tr	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	TC	AA	CA	TT	AA	AA	CC	CC	GG	CA	AT	GA	TT	GA	GG	TC	AA	GA	TT	TC	TT	TT	TT	CC	CC	CC	GG	GG	AA
30	Tr	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CA	TT	GA	AA	CC	CC	GG	CA	AA	AA	TT	AA	AA	CC	AA	GG	TT	TT	TT	TT	TT	CC	TT	CC	GG	GG	AA
30	Tr	CC	TT	TT	GG	TT	CC	CC	TT	TT	TT	CC	AA	CC	TT	AA	AA	CC	CC	GG	CC	AA	AA	TT	AA	AA	CC	AA	GG	TT	TT	TT	TT	TT	CC	TT	CC	GG	AA	AA
25	Tr	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CC	TT	AA	AA	CC	CC	GG	CC	AA	AA	TT	AA	AA	CC	AA	GG	TT	TT	TT	TT	TT	CC	TT	CC	GG	AA	CC
25	Tr	CC	TT	AA	GG	TT	CC	TT	AA	TT	TT	CC	GA	AA	TT	AA	AA	CC	CC	GG	AA	TT	GG	TT	GG	GG	TT	AA	AA	TT	CC	TT	TT	TC	CC	TT	CC	GG	GG	AA

Table A 2 continued. *M. edulis* genotypes. The first column (L%) shows the percentage of neoplastic-looking cells in the hemolymph sample. The second column indicates the phenotype groups defined in Table 4.2: L-late leukemic, shaded, Tr – transitional, N- normal.

L%	phen	snp158	snp182	snp191	snp204	snp219	snp231	snp261	snp262	snp267	snp297	snp347	snp380	snp386	snp392	snp401	snp407	snp416	snp419	snp428	snp465	snp524	snp539	snp543	snp545	snp552	snp560	snp563	snp566	snp575	snp605	snp680	snp731	snp737	snp740	snp767	snp821	snp854	snp914	snp917
23	Tr	CC	TT	TT	GA	TT	TC	TT	AT	TG	TT	TC	GA	CA	TG	AA	AT	CA	CA	GG	CA	AA	AA	TT	AA	GA	CC	AA	GG	AT	TT	TT	TT	TC	CC	CC	CC	GA	GG	AA
20	Tr	CC	TT	AT	GA	TT	CC	TT	AA	TT	TT	CC	AA	CA	TT	AA	AA	CC	CC	GG	CC	AA	AA	TT	AA	GA	CC	AA	GG	TT	TT	TT	TT	TT	CC	CC	CC	GG	CC	CC
12	Tr	TC	TT	AT	GA	TT	CC	TT	AT	TG	TC	TC	GA	AA	TG	GA	AA	CC	CA	GG	AA	AT	GA	TT	GA	GA	TC	AA	GG	AT	TC	TC	TC	TC	CA	CC	TC	GA	AA	CC
10	Tr	CC	TT	AT	GG	TT	CC	CC	TT	TT	TT	CC	AA	AA	TT	GA	AA	CC	CC	GG	AA	AT	GA	TT	GA	GA	TC	AA	GA	TT	TC	TT	TT	TT	CC	TT	CC	GG	GG	AA
8	N	CC	TT	AT	GG	TT	CC	TT	AA	TT	TT	CC	AA	AA	TT	AA	AA	CC	CC	GG	CA	AT	GA	TT	GA	GG	TC	AA	GA	TT	TT	TT	TT	TT	CC	TT	CC	GG	AA	AA
7	N	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CC	TT	AA	AA	CC	CC	GG	CC	AA	AA	TT	AA	AA	CC	AA	GG	TT	TT	TT	TT	TT	CC	CC	CC	GG	CC	CC
6	N	CC	TT	TT	GA	TT	CC	TT	AA	TT	TT	TC	AA	CA	TG	GA	AT	CC	CA	GG	CA	AA	AA	TT	AA	AA	CC	AA	GG	TT	TT	TT	TT	TT	CC	CC	CC	GG	GG	AA
5	N	CC	TT	AT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CA	TT	AA	AA	CC	CC	GG	CA	AT	GA	TT	GA	GA	TC	AA	GA	TT	TC	TT	TT	TT	CC	TT	CC	GG	GG	AA
5	N	CC	TT	TT	GG	TT	CC	CC	TT	TT	TT	TC	AA	AA	TT	AA	AA	CC	CC	GG	AA	AA	AA	TC	AA	GG	CC	GA	GG	TT	TT	TT	TT	TT	CC	TC	CC	GG	GG	AA
3	N	CC	TT	AT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CA	TT	AA	AA	CC	CC	GG	CA	AT	GA	TT	GA	GA	TC	AA	GA	TT	TC	TT	TT	TT	CC	TT	CC	GG	CC	CC
2	N	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CC	TT	AA	AA	CC	CC	GG	CC	AA	AA	TT	AA	AA	CC	AA	GG	TT	TT	TT	TT	TT	CC	TT	CC	GG	AA	AA
2	N	CC	TT	AT	GG	TT	CC	CC	TT	TT	TT	CC	AA	AA	TT	AA	AA	CC	CC	GG	AA	TT	GG	TT	GG	GG	TT	AA	AA	TT	CC	TT	TT	TT	CC	TT	CC	GG	AA	AA
1	N	CC	TT	AA	GG	TT	CC	TT	AA	TT	TT	CC	AA	AA	TT	AA	AA	CC	CC	GG	AA	AT	GA	TC	GA	GA	TC	AA	GA	TT	TT	TT	TT	TT	CC	TC	CC	GG	GG	AA
1	N	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CA	TT	GA	AA	CC	CC	GG	CA	AA	AA	TT	AA	GA	CC	AA	GG	TT	TT	TT	TT	TT	CC	CC	CC	GG	AA	AA
1	N	CC	TT	AT	GA	TT	CC	TT	AA	TT	TT	CC	AA	CA	TT	AA	AA	CC	CC	GG	CA	AT	GA	TT	GA	GA	TC	AA	GA	TT	TC	TT	TT	TT	CC	CC	CC	GG	GG	AA
1	N	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CA	TT	AA	AA	CC	CC	GG	CC	AA	AA	TT	AA	GA	CC	AA	GG	TT	TT	TT	TT	TT	CC	CC	CC	GG	GG	AA
1	N	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CA	TT	GA	AA	CC	CC	GG	CA	AA	AA	TT	AA	AA	CC	AA	GG	TT	TT	TT	TT	TT	CC	CC	CC	GG	GG	AA
1	N	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CC	TT	AA	AA	CC	CC	GG	CC	AA	AA	TT	AA	AA	CC	AA	GG	TT	TT	TT	TT	TT	CC	CC	CC	GG	GG	AA
1	N	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CC	TT	AA	AA	CC	CC	GG	CC	AA	AA	TT	AA	AA	CC	AA	GG	TT	TT	TT	TT	TT	CC	CC	CC	GG	AA	CC
1	N	CC	TT	TT	GG	TT	CC	CC	TT	TT	TT	TC	AA	CA	TT	AA	AA	CC	CC	GG	CA	AA	AA	TC	AA	GA	CC	GA	GG	TT	TT	TT	TT	TT	CC	TC	CC	GG	GG	AA
1	N	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CC	TT	AA	AA	CC	CC	GG	CC	AA	AA	TT	AA	AA	CC	AA	GG	TT	TT	TT	TT	TT	CC	CC	CC	GG	GG	AA
1	N	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	GA	CC	TT	AA	AA	CC	CC	GG	CC	AA	AA	TT	AA	AA	CC	AA	GG	TT	TT	TT	TT	TT	CC	TT	CC	GG	CC	CC
0	N	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CC	TT	AA	AA	CC	CC	GG	CC	AT	GG	TT	GA	GG	TT	AA	AA	TT	TC	TT	TT	TT	CC	CC	CC	GG	GG	AA
0	N	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CC	TT	AA	AA	CC	CC	GG	CA	AA	AA	TC	AA	GG	CC	GA	GG	TT	TT	TT	TT	TT	CC	TC	CC	GG	GG	AA
0	N	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CC	TT	AA	AA	CC	CC	GG	CA	AT	GA	TC	GA	GG	TC	AA	GA	TT	TC	TT	TT	TT	CC	TC	CC	GG	GG	AA

Table A 3 continued. *M. edulis* genotypes. The first column (L%) shows the percentage of neoplastic-looking cells in the hemolymph sample. The second column indicates the phenotype groups defined in Table 4.2: L-late leukemic, shaded, Tr – transitional, N- normal.

0	N	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	AA	TT	AA	AA	CC	CC	GG	AA	TT	GG	TT	GG	GG	CC	AA	AA	TT	TT	TT	TT	TT	CC	CC	CC	GG	AA	AA
0	N	CC	TT	TT	GG	TT	CC	TT	AA	TT	TT	CC	AA	CA	TT	AA	AA	CC	CC	GG	CA	AA	AA	TT	AA	GA	CC	AA	GG	TT	TT	TT	TT	TT	CC	CC	CC	GG	AA	AA
0	N	CC	TT	AT	GA	TT	TC	TT	AA	TT	TT	TC	AA	AA	TT	AA	AA	CC	CC	GG	AA	AA	GA	TC	GA	GG	TC	AA	GA	TT	TT	TT	TT	TT	CC	TC	CC	GG	CC	CC
0	N	CC	TT	AT	GA	TT	CC	TT	AA	TT	TT	CC	AA	CC	TT	AA	AA	CC	CC	GG	CC	AA	AA	TT	AA	GA	CC	AA	GG	TT	TT	TT	TT	TT	CC	CC	CC	GG	CC	CC

Table A 4. Theoretically predicted haplotypes for *M. edulis*; haplotype total frequencies; counts and frequencies of haplotypes in case group (diseased, >10% of neoplastic cells in hemolymph) and control group (non-diseased, <10% of neoplastic cells in hemolymph), Chi square and p value for association of the haplotypes with the disease.

	Halotypes	Total freq	Case counts	Contr counts	Case Ratio	Contr Ratio	chi squ	p value
1	CTTGTCTATTCACTAACCGCAATAACAGTTTTTCCCGGA	0.185	9.0 : 53.0,	12.4 : 39.6	0.145	0.239	1.629	0.2018
2	CTTGTCTATTCACTAACCGCAATAACAGTTTTTCTCGGA	0.078	6.0 : 56.0,	3.0 : 49.0	0.097	0.058	0.594	0.4411
3	CTTGTCTATTCACTAACCGCAATAGCAGTTTTTCCCGGA	0.046	2.0 : 60.0,	3.3 : 48.7	0.032	0.064	0.642	0.4229
4	CTTGTCTATTCAATAACCGATGTGGTAATTTTTCTCGGA	0.039	3.0 : 59.0,	1.5 : 50.5	0.048	0.029	0.289	0.5905
5	CTTGTCTATTCAATAACCGCAATAGCAGTTTTTCCCGGA	0.037	3.0 : 59.0,	1.2 : 50.7	0.048	0.024	0.469	0.4934
6	CTAGTCTATTCAATAACCGATGTGGTAATCTTTCTCGGA	0.035	1.0 : 61.0,	3.0 : 49.0	0.016	0.058	1.455	0.2277
7	TCTACCCTGCTGAGATAACAAATAGCAGATCCCACCATC	0.022	2.5 : 59.5,	0.0 : 52.0	0.04	0	2.143	0.1432
8	TCTACCCTGCTGAGATAACAAATAGCAGATCCCACTATC	0.022	2.5 : 59.5,	0.0 : 52.0	0.04	0	2.143	0.1432
9	CTTGTCTATTCAATAACCGATGTGGCAATTTTTCCCGGA	0.017	0.0 : 62.0,	2.0 : 50.0	0	0.038	2.428	0.1192
10	CTTGTCCTTTCACTAACCGCAATAACAGTTTTTCTCGGA	0.017	2.0 : 60.0,	0.0 : 52.0	0.032	0	1.707	0.1914
11	CTTGTCTATTCAATAACCGAAATAGCAGTTTTTCTCGGC	0.017	1.0 : 61.0,	1.0 : 51.0	0.016	0.019	0.016	0.8999
12	CTTGTCTATTCACTAACCGCAATAACAGTTTTTCTCGGC	0.017	0.0 : 62.0,	2.0 : 50.0	0	0.038	2.428	0.1192
13	CTTGTCTATTCAATGACCGAAATAACAGTTTTTCTCGGA	0.017	1.0 : 61.0,	1.0 : 51.0	0.016	0.019	0.016	0.8999
14	CTTGTCTATTCAATAACCGAAATAGCAGTTTTTCTCGGA	0.017	2.0 : 60.0,	0.0 : 52.0	0.032	0	1.707	0.1914
15	CTTACCCTGCTGAGATAACAAATAGCAGATCCCACCATC	0.013	1.5 : 60.5,	0.0 : 52.0	0.024	0	1.275	0.2589
16	CTTACCCTGCTGAGATAACAAATAGCAGATCCCACTATC	0.013	1.5 : 60.5,	0.0 : 52.0	0.024	0	1.275	0.2589
17	CTTGTCTATTCAATAACCGATGTGGTAATCTTTCTCGGA	0.013	0.0 : 62.0,	1.5 : 50.5	0	0.029	1.808	0.1787
18	CTAGTCTATTCAATAACCGCAATAGCAGTTTTTCTCGGA	0.013	0.0 : 62.0,	1.5 : 50.5	0	0.029	1.808	0.1787
19	CTTGTCTATTCAATGACCGAAATAACAGTTTTTCCCGGA	0.011	0.0 : 62.0,	1.3 : 50.7	0	0.026	1.606	0.205
20	CTTGTCTATTCACTAACCGAAATAACAGTTTTTCCCGGA	0.011	1.0 : 61.0,	0.2 : 51.7	0.016	0.005	0.338	0.561